

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 14:39:25 ; Search time 3772.55 Seconds  
(without alignments)  
10929.928 Million cell updates/sec

Title: US-09-806-276A-4  
Perfect score: 2546  
Sequence: 1 cggggctagcccgagagacc.....cagctanaacgagggcagta 2546

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	882.8	34.7	1032	13	BM458203
2	879.2	34.5	930	9	AL545294
3	872.4	34.3	938	14	BM804771
4	854.6	33.6	920	9	AL547013
5	850.4	33.4	899	9	AL520266
6	844.6	33.2	1968	11	AK019447

7	844	33.2	1970	11	AK014281
8	842.2	33.1	887	9	AL572131
9	818	32.1	852	9	AL545262
10	795.8	31.3	889	13	BI916234
11	789.8	31.0	870	13	BI760499
12	783.6	30.8	827	13	BI918195
13	783	30.8	869	14	BQ680912
14	774	30.4	786	13	BI915037
15	770.2	30.3	959	14	BQ878634
16	767.8	30.2	925	13	BI906659
17	763.6	30.0	1487	11	AK012309
18	761	29.9	918	13	BI758178
19	751.8	29.5	866	13	BI861738
20	748.4	29.4	773	13	BI908681
21	744.6	29.2	1031	14	BQ067953
22	725	28.5	808	13	BI767600
23	724.6	28.5	879	14	BQ224749
24	708	27.8	920	14	BQ961461
25	702	27.6	720	14	BM981107
26	701	27.5	876	14	BQ435221
27	697.8	27.4	705	12	BG703895
28	678.6	26.7	1160	12	BG036153
29	671.4	26.4	952	13	BI912055
30	668.8	26.3	704	13	BI760730
31	662.8	26.0	1123	12	BG386157
32	659.4	25.9	685	14	BM990899
33	655.4	25.7	664	13	BI765476
34	649.2	25.5	750	12	BG750650
35	646.8	25.4	1122	14	BQ055733
36	642.4	25.2	1031	12	BG027198
37	636.8	25.0	703	12	BG722248
38	635.8	25.0	927	14	BQ228776
39	630	24.7	631	10	BE385516
40	626.4	24.6	959	12	BG122190
41	625.4	24.6	744	13	BI912364
42	625	24.5	909	12	BE889390
43	623.8	24.5	811	12	BF305603
44	613.8	24.1	700	9	AL544300
45	610.8	24.0	646	10	BE206781

ALIGNMENTS

RESULT 1  
BM458203  
LOCUS  
DEFINITION  
AGENCOURT\_6411553 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5530576  
1032 bp mRNA linear EST 05-FEB-2002  
5', mRNA sequence.  
ACCESSION  
BM458203  
VERSION  
BM458203.1 GI:18507243  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1032)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12210 row: n column: 17  
High quality sequence stop: 629.  
FEATURES  
Location/Qualifiers  
1..1032  
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5530576"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 298 a 187 c 200 g 347 t
ORIGIN
Query Match 34.7%; Score 882.8; DB 13; Length 1032;
Best Local Similarity 98.6%; Pred. No. 1.5e-169;
Matches 901; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
QY 1079 TCCCTCTCTGACCTTACAGGTTCCCTCTCCCTGCAGCTTGATCTTGGGCCCTCACTT 1138
Db 1 TCCCTCTCTGACCTTACAGGTTCCCTCTCCCTGCAGCTTGATCTTGGGCCCTCACTT 60
QY 1139 TTGGATGAGGTTGCTGAATGTAATGGATAAAATAAGTAACAAGATGCCAACTTTTTCCT 1198
Db 61 TTGGATGAGGTTGCTGAATGTAATGGATAAAATAAGTAACAAGATGCCAACTTTTTCCT 120
QY 1199 TTGGGGTAAAGGTACAAAACAACTAACACAGTTGAAGAGAAGGGCTTCCGGAGCTG 1258
Db 121 TTGGGGTAAAGGTACAAAACAACTAACACAGTTGAAGAGAAGGGCTTCCGGAGCTG 180
QY 1259 TATTGTCAGTTTGTGTTGGGTTTCTTAAATAATAATCTTACAAAAGTATTTTTCACCT 1318
Db 181 TATTGTCAGTTTGTGTTGGGTTTCTTAAATAATAATCTTACAAAAGTATTTTTCACCT 240
QY 1319 GTTATGCCCTGTTTGCAAAACAAATTTAGAAAAAACAACAAAGCAAAACCTATCTTGGC 1378
Db 241 GTTATGCCCTGTTTGCAAAACAAATTTAGAAAAAACAACAAAGCAAAACCTATCTTGGC 300
QY 1379 AAAAAAGGAAGTGTGAGTCAGAGCCCATTTTCAGGAGGCAATTTGGTGATGTTTCGGCTCACAT 1438
Db 301 AAAAAAGGAAGTGTGAGTCAGAGCCCATTTTCAGGAGGCAATTTGGTGATGTTTCGGCTCACAT 360
QY 1439 ATTGTTTGCAGACACACAAGAAATCTGGCTTGGCCAGGATTTGGCACTAGCTATGAAGGGC 1498
Db 361 ATTGTTTGCAGACACACAAGAAATCTGGCTTGGCCAGGATTTGGCACTAGCTATGAAGGGC 420
QY 1499 TGAGCGAGTCACATTAAGGAACCTTCACGGAACCTTTATAGCACTCCGACATTTTCGAGCA 1558
Db 421 TGAGCGAGTCACATTAAGGAACCTTCACGGAACCTTTATAGCACTCCGACATTTTCGAGCA 480
QY 1559 AGAGGAAGTCAAAATTTATTTAACACCTAAGCCCTTTTGTAGACTCTTTTCTATATATG 1618
Db 481 AGAGGAAGTCAAAATTTATTTAACACCTAAGCCCTTTTGTAGACTCTTTTCTATATATG 540
QY 1619 CTTAGGCTCACCATAGCGAATTCGCCAGTGTAAAAACTTTTCTGTTTTCACATTTGAAC 1678
Db 541 CTTAGGCTCACCATAGTGAATTCGCCAGTGTAAAAACTTTTCTGTTTTCACATTTGAAC 600
QY 1679 TTATGGGTTTGGGGATTCTTCTGTAGTTCTTATATATATATATATATATATATATAT 1738
Db 601 TTATGGGTTTGGGGATTCTTCTGTAGTTCTTATATATATATATATATATATATATAT 660
QY 1739 GCAAAATTTTGACTGTCAGCTACATGTTGGTAAGACACAGGCAAAAGTATTACTGTAACTA 1798
Db 661 GCAAAATTTTGACTGTCAGCTACATGTTGGTAAGACACAGGCAAAAGTATTACTGTAACTA 720
QY 1799 AGTTATTTTAAAGTTAAATATATATTTTACG-TGCCCTTGGCTTTTATTTGCAGAGTCT 1857
Db 721 AGTTATTTTAAAGTTAAATATATATTTTACGCTGCCCTTGGCTTTTATTTGCAGAGTCT 780
QY 1858 ACATTTTATAGATTCTACATCAGATGTTGTCACTTATTTCCATTTGGGATTTCCATTTGAAG 1917
Db 781 ACATTTTATAGATTCTACATCAGATGTTGTCACTTATTTCCATTTGGGATTTCCATTTGAAG 840
QY 1918 CTGTGTATGTGCGGTGTTTGGAAAAAGTGATTATTCATACCTTAGTTTCTTTTCTTCTCATCTGTT 1977
```

```
Db 841 CTGGGTATGGCGGTGTGCAAAAGGGTATTATACACTTAGTTTCTTCTTCTCATCCGGT 900
QY 1978 ATCATACTTTTAAAC 1991
Db 901 ATCCAACCTTTTAC 914
RESULT 2
AL545294 298 a 187 c 200 g 347 t
LOCUS AL545294 LTI_NFL006_PL2 930 bp mRNA linear EST 16-FEB-2001
DEFINITION AL545294 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI028YL18 5
prime, mRNA sequence.
ACCESSION AL545294
VERSION AL545294.1 GI:12877775
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. .930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI028YL18"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 226 a 268 c 240 g 193 t 3 others
ORIGIN
Query Match 34.5%; Score 879.2; DB 9; Length 930;
Best Local Similarity 99.6%; Pred. No. 8.2e-169;
Matches 889; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
QY 1 CGGGGCTAGCCCGGAGACCCCGCCACCGCGCTGGGGCGCTTCAGCCCGTCTCGGAGCGG 60
Db 39 CGNGGCTAGCCCGGAGACCCCGCCACCGCGCTGGGGCGCTTCAGCCCGTCTCGGAGCGG 98
QY 61 ATAATGCGGTGACGAGGACCAACCGCCGCGAGACTCGGCTGGATCTGCGCACAGCGGCAGG 120
Db 99 ATAATGCGGTGACGAGGACCAACCGCCGCGAGACTCGGCTGGATCTGCGCACAGCGGCAGG 158
QY 121 GATTGCGTGCCTCCCGGGAGGCGCCGCGGAGAGTCTGAGTAACTGGCTTCTAGCGGCGCGGTT 180
Db 159 GATTGCGTGCCTCCCGGGAGGCGCCGCGGAGAGTCTGAGTAACTGGCTTCTAGCGGCGCGGTT 218
QY 181 TGTCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCCTCACTTGGC 240
Db 219 TGTCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCCTCACTTGGC 278
QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 300
Db 279 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 338
QY 301 TTGAGAAAAATGATGTGACAGTTCCCTATCAAAAAGGATTCAGAAACATATACCATCTGTG 360
```

||||| 339 TTGAGAAAAATGATGTGACAGTTCTCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 398  
QY ||||| 361 AAGAAAGTGGCCCTTCTCCCGCTTGCAAAATAGACATTCTCAATTCCAAATGCCAGC 420  
Db ||||| 399 AAGAAAGTGGCCCTTCTCCCGCTTGCAAAATAGACATTCTCAATTCCAAATGCCAGC 458  
QY ||||| 421 CAAGACCCCAATTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTTAAACTGAG 480  
Db ||||| 459 CAAGACCCCAATTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTTAAACTGAG 518  
QY ||||| 481 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGAGAGACTTTCGCCACACCATCCA 540  
Db ||||| 519 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGAGAGACTTTCGCCACACCATCCA 578  
QY ||||| 541 CATTTGCAAAAGAGGCGCAGCAGATGTCTTTGGAGATATTTCTTTCTCAAGGGAACCTA 600  
Db ||||| 579 CATTTGCAAAAGAGGCGCAGCAGATGTCTTTGGAGATATTTCTTTCTCAAGGGAACCTA 638  
QY ||||| 601 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 660  
Db ||||| 639 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 698  
QY ||||| 661 GTTCTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAGAAACGCCCTCCCGGTGCT 720  
Db ||||| 699 GTTCTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAGAAACGCCCTCCCGGTGCT 758  
QY ||||| 721 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTATT 780  
Db ||||| 759 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTATT 818  
QY ||||| 781 GTCACAGTGACATTTAATTCAAACAGGAGTCTCGGGCCAGCAAAAGTGGCCAGGCT 840  
Db ||||| 819 GTCACAGTGAMATTTAATTCAAACAGGAGTCTCGGGCCAGCAAAAGTGGCCAGGCT 878  
QY ||||| 841 TAGCTGCGAGCCCGTCATGAGGAGAAAGCTCAGAGAGAAAGCAGTCTGTGG 893  
Db ||||| 879 TAGCTGCGAGCCCGTCATGAGGAGAAAGCTCAGAGAGAAAGCAGTCTGTGG 930

RESULT 3  
BM804771  
LOCUS BM804771 938 bp mRNA linear EST 05-MAR-2002  
DEFINITION AGENCOURT\_6495033 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5725457  
5', mRNA sequence.

ACCESSION BM804771  
VERSION BM804771.1 GI:19121594  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 938)

REFERENCE NIH-MGC <http://mgi.ncl.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Invitrogen

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12716 row: f column: 18  
High quality sequence stop: 659.

FEATURES  
Location/Qualifiers  
1..938

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5725457"  
/clone\_lib="NIH\_MGC\_125"

/lab\_host="DH10B"  
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;  
Site\_1: EcoRV (destroyed); Site\_2: NotI; RNA source pool  
of three ovaries, from females ranging in age from 38 to  
49 yo. Library is oligo-dT primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 2.1 kb, insert size range 1-3.5 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 036."

BASE COUNT 230 a 269 c 249 g 189 t 1 others  
ORIGIN

Query Match 34.3%; Score 872.4; DB 14; Length 938;  
Best Local Similarity 98.5%; Pred. No. 2e-167;  
Matches 901; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTCACGCCGCTCTCGGAGCGG 60  
Db ||||| 25 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTCACGCCGCTCTCGGAGCGG 84  
QY 61 ATATGCGGTGAGCAGGCACACCGCCGCGAGACTCGGTGGATCTGCGCACAGCGCAGG 120  
Db ||||| 85 ATATGCGGTGAGCAGGCACACCGCCGCGAGACTCGGTGGATCTGCGCACAGCGCAGG 144  
QY 121 GATTGCGTGCCTCCCGGGAGGCGCGGCGGCGGCTGGGATCCTCAGCGCGCGCGGTT 180  
Db ||||| 145 GATTGCGTGCCTCCCGGGAGGCGCGGCGGCGGCTGGGATCCTCAGCGCGCGCGGTT 204  
QY 181 TGTCCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCACTGGC 240  
Db ||||| 205 TGTCCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCACTGGC 264  
QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 300  
Db ||||| 265 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 324  
QY 301 TTGAGAAAAATGATGTGACAGTTCTCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 360  
Db ||||| 325 TTGAGAAAAATGATGTGACAGTTCTCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 384  
QY 361 AAGAAAGTGGCCCTTCTCCCGCTTGCAAAAATAGACATTCTCAAAATTCCAAATGCCAGC 420  
Db ||||| 385 AAGAAAGTGGCCCTTCTCCCGCTTGCAAAAATAGACATTCTCAAAATTCCAAATGCCAGC 444  
QY 421 CAAGACCCCAATTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTTAAACTGAG 480  
Db ||||| 445 CAAGACCCCAATTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTTAAACTGAG 504  
QY 481 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGAGAGACTTTCGCCACACCATCCA 540  
Db ||||| 505 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGAGAGACTTTCGCCACACCATCCA 564  
QY 541 CATTTGCAAAAGAGGCGCAGCAGATGTCTTTGGAGATATTTCTTTCTCAAGGGAACCTA 600  
Db ||||| 565 CATTTGCAAAAGAGGCGCAGCAGATGTCTTTGGAGATATTTCTTTCTCAAGGGAACCTA 624  
QY 601 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 660  
Db ||||| 625 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 684  
QY 661 GTTCTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAGAAACGCCCTCCCGGTGCT 720  
Db ||||| 685 GTTCTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAGAAACGCCCTCCCGGTGCT 744  
QY 721 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTATT 780  
Db ||||| 745 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTATT 804  
QY 781 GTCACAGTGACATTTAATTCAAACAGGAGTCTTCGGGGCAGCAAAAGTGGCCAGGCT 840  
Db ||||| 805 GTCACAGTGACATTTAA-TCCCAACAGGAGTCTTCGGGGCAGCAAAAGTGGCCAGGCT 863















GAGAGAGAGATTCTCGAGTTAATTAATTAATCCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.

FEATURES

Source Location/Qualifiers  
1..1970  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:3200001F04"  
/db\_xref="MGD:MGI:1898924"  
/db\_xref="taxon:10090"  
/clone="3200001F04"  
/tissue\_type="head"  
/dev\_stage="14, 17 days embryo"  
505..1269  
/note="data source:SPTR, source key:Q9UKI2, evidence:ISS homolog to CRIB-CONTAINING BORG2 PROTEIN putative"  
/codon\_start=1  
/protein\_id="BAB29241.1"  
/db\_xref="GI:12852023"  
/db\_xref="MGD:MGI:1915275"  
/translation="MPAKTPIYLKAANNKKGKFKLRDILSPDMISPLGLDFRHTIHI  
KGEQHDVDFISLQGNLYELLPGNQEKAHSGQFPGHNDFFRANSTSDSMFTETPSV  
LKNALSLPTIGGSQALMLPLSPVTFHKSQESFGRPKLPRLSCPEVMEKVKVQESLL  
ENGAHVQGDTSWSSGSGSSQSSQGRDSSSLSEQSSDWPADDMFEHPASCELVSKT  
KSEESFSDLTGLSLQLDLGPSLLDEVLMVDKNK"

CDS

BASE COUNT	517 a	490 c	467 g	496 t
ORIGIN				
Query Match	33.2%; Score 844; DB 11; Length 1970;			
Best Local Similarity	70.8%; Pred. No. 8.8e-162;			
Matches 1325; Conservative	0; Mismatches 465; Indels 82; Gaps 12;			
QY	20	CGGCCACCGGCTGGGGCGCTTACGCCGCTCGGAGCGGATATGCGGTGAGCAGGCA	79	
Db	109	CGGCTGCTCCCGGGAGAGCGCGCGCAGTCCCGGAGCAGATTCTGCGGTGAACCGTTG	168	
QY	80	CCAGCCCGGACAGCTCGGCTGGATCTGGC-----CACAGCGGCAGGGATTGCGT	128	
Db	169	CGGCTCTGCTTCTGGCGCACCGAGCGGAGTGTTCGCGCTCGGAGCGGACCGG	228	
QY	129	CGCCCGCGGGAGCGCGGGGAGCGGCTCGGATCTCTAGGAAGCTCACTTGGCCGCTGG	188	
Db	229	GCCGACGCGCGGCTCCCGGAGAGCGCGCGCAGTCTAGAGCTGCTGTTCTCTGAG	288	
QY	189	TGTGGTCAAGACTGGATGATGTAACCTGCTCTAGGAAGCTCACTTGGCCGCTAACCT	248	
Db	289	CTGTGAAGGACTCGATAATGTAACCTGCTCTCGGAAAGACCGCGTGGCCCTGCGCT	348	
QY	249	CAGGAAGGTTCTCTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCACTTGAGAA	308	
Db	349	GAGCAAGATCTCCAAGACCCCTGCTGCTTTTAGAGCCACTTGAGAAG-----A	397	
QY	309	AATGATGTGACAGTTCTATCAAAAAGGA--TTCAGAAACATATACCATCTGTGAAGAA	365	
Db	398	AATGATGTGACAAATCTTATGAAAAGGACGTTTGGAGTGTAGGATACCGGTGCAGAA	457	
QY	366	AGTGCCCTTCTCTCCGCTTGCAAAATAGACATTTCAAAATCCAAATCCAAATGCCTGCCA	425	
Db	458	GGTCACTCTTCCCTGGTCTGTGAAATAGACATTTCAAAATCCAAATCCAAATGCCTGCCA	517	
QY	426	CCCCAATTTACCTGAAAGCAGCAATAACAAGAAAGGAAAGAAATTTAACTGAGGGACA	485	
Db	518	CCCCAATTTACCTGAAAGCGGTAAACAACAAGAAAGGAAAGAAATTTAACTGAGGGACA	577	
QY	486	TTCTGTCTCTGATATGATGATCAGTCCCGCTTGGAGACTTTCCGACACCATCCACATTG	545	
Db	578	TTCTGTCTCTGATATGATGATCAGTCCCGCTTGGAGACTTTCCGACACCATCCACATTG	637	
QY	546	GCAAAGAGGGCCAGCAGGATGCTCTTTGGAGATATTTCTTTTCAAGGGAACCTACGAGC	605	
Db	638	GCAAAGAGGGCCAGCAGGATGCTCTTTGGGAGATATTTCTTTTCTTCAAGGGAATTA	697	

QY	606	TTTTACTCTGAAACACGAGGAGAAAGCACACCTGGGCGAGTTCCCTGGGCAATAATGAGTTCT	665	
Db	698	TCTTGGCGGAAACCAAGAGAGGACACACTCTGGCCAGTTCCCGGGCATAACGACTTCT	757	
QY	666	TCCGGGCAACAGCACCTCGGACTCTGTGTTTACAGAAAACGCCCTCCCGGTGCTCAAAA	725	
Db	758	TCCGGGCAACAGCACCTCGGACTCAATGTTTACAGAAAACGCCCTCCCGGTGCTCAAAA	817	
QY	726	ATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATTTGTCAC	785	
Db	818	ATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATTTGTCAC	877	
QY	786	CAGTGACATTTAATTCAAAACAGGAGTCTTCGGGCCAGCAAAAGCTGCCAGGCTTAGCT	845	
Db	878	CAGTGACATTTAATTCAAAACAGGAGTCTTCGGGCCAGCAAAAGCTGCCAGGCTTAGCT	937	
QY	846	GCGAGCCCGTCAATGGAGGAAAAAGCTCAGAGAAAAAGCAGTCTGTTGGAGAAATGGGACAG	905	
Db	938	GTGAGCCTGTCAATGGAGGAAAAAGCTCAGAGAAAAAGCAGTCTGTTGGAGAAATGGGACAG	997	
QY	906	TCCACCAGGAGACACCTCGTGGGCTCCAGCGGTTCTGCATCTCAGTCCAGCCCAAGGCA	965	
Db	998	TCCACCAGGAGACACCTCGTGGGCTCCAGCGGTTCTGCATCTCAGTCCAGCCCAAGGCC	1057	
QY	966	GAGACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGAGTGGCCAGCCGAGACATGT	1025	
Db	1058	GCGACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGAGTGGCCAGCCGAGACATGT	1117	
QY	1026	TTGACCATCCACCCCATCGGAGCTCATCAAGGAAAGCAGTAAAGTCAAGAGTCCCTCT	1085	
Db	1118	TTGAGCATCTCTGCTCTGAGCTGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1177	
QY	1086	CTGACCTTACAGGTTCCCTCCCTCTCCCTGCAGCTTGTGTTGGCCCTCAGTTTGGATG	1145	
Db	1178	CTGACCTGACGGCTCCCTGCTCTCCCTGCAGTTGGATCTTGGCCCTTCACTTTGGATG	1237	
QY	1146	AGTGCTGAATGTAATGGATAAAATAAGTAACAAGTCCCAACTTTTTCCTTTGGGGT	1205	
Db	1238	AAGTCTGAATGTCATGGATAAAATAAGTACCCAGAGTCAAGTCAAGTCAAGTCAAGT	1297	
QY	1206	AAAAGGTAC-----AAAAACAACTAACACACAGTGAAGAGAGGG	1246	
Db	1298	AAGAGGTACCAAAATAAGCAGAAAAACAAAAATAAATCAACACAGCTGATGAAAGAG	1357	
QY	1247	CTTCCGAGCTG-----TATTTGCAGTTTGTGTGGGTTTCTAAAAATAA-----TATT	1296	
Db	1358	CTGCTCTGACTGTTTATTTGCAGCCATGTGATGCTATTTCTAAAAATAAAGCTCTCTGT	1417	
QY	1297	CTTACAAAAGTATTTTTCCTGTTATGCTGCTGTTTGCAAA-----AACAAAT	1344	
Db	1418	CTTACAAAAGTATTTTTCCTGTTATGCTGCTGTTTGCAAA-----AACAAAT	1477	
QY	1345	TAGAAAAACAAACAAAGCAAAACCTATCTTGGCAAAAAAGGAAGTGAAGTCAAGAGCCCA	1404	
Db	1478	TTTAAAAATTAATAAAACAAACAAACAAACAAACAAACCTGTTCTGCAAAAAAGAAAAA	1537	
QY	1405	TTTTCAGGAGGCAATGGTGTGTTTCGGCTCACATATTTGTTG--CAGACACACAAGAAAT	1462	
Db	1538	AAAGAAAAAGAAAAAGGAGGAGTGAAGTGAAGGCTGAGCCAGTCAATTAAGGAACTT	1597	
QY	1463	CTGGCTTGGCCAGGATTTGGCACTAGCTATGAAGGCTGAGCCAGTCAATTAAGGAACTT	1522	
Db	1598	CATACATGCTCAGTTTGTGTTGTAGACATGCAGAAATCCAGGCTGTCTAAGAGGCTGGG	1657	
QY	1523	CACGGAACTTTATAGCACTCCGACATTTTCTGAGCAAGAGGAAGTCA-AAATTTATTAA	1581	
Db	1658	TCCATGGGTCTCTCCAAGGCAATTCATTCATGCTGTTATGAGAAGGAAACTTACTTATGG	1717	
QY	1582	CACCTAAGCCCTTTTGTAGACTCTTTCT-----ATATATTGCTTAGGCTCACCATAGC	1635	
Db	1718	CACCTAAGCCCTTTTTCCTAGCCCTTTTGTATATTTGTATATTTGGGTTTCAACACAGC	1777	



```
/clone="CSODI028YL18"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      291 a      155 c      142 g      255 t      9 others
ORIGIN

Query Match      32.1%; Score 818; DB 9; Length 852;
Best Local Similarity 98.2%; Pred. No. 2.4e-156;
Matches 820; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1278 GGTTCCTAAATAATATCTTACAAAGTATTTTTTACCTGTTATGCCCTGTTTGCAAA 1337
Db ||:|||||
QY 852 GGTTCCTAAATAATATCTTACAAAGTATTTTTTACCTGTTATGCCCTGTTTGCAAA 793
Db ||:|||||
QY 1338 AACAAATTTAGAAAAAACAAACAAAGCAAAACCTATCTTGGCAAAAAAGGAAGTGAATCA 1397
Db ||:|||||
QY 792 AACAAATTTAGAAAAAACAAACAAAGCAAAACCTATCTTGGCAAAAAAGGAAGTGAATCA 733
Db ||:|||||
QY 1398 GAGCCCATTTTCAGGAGGCATTGGTGATGTTTCGGCTCACATATGTTTGCAGACACAAA 1457
Db ||:|||||
QY 732 GAGCCCATTTTCAGGAGGCATTGGTGATGTTTCGGCTCACATATGTTTGCAGACACAAA 673
Db ||:|||||
QY 1458 GAAATCTGGCTGGCCAGGATTGGCACTAGCTATGAAGGCTGAGCGAGTCACATTAAGG 1517
Db ||:|||||
QY 672 GAAATCTGGCTGGCCAGGATTGGCACTAGCTATGAAGGCTGAGCGAGTCACATTAAGG 613
Db ||:|||||
QY 1518 AACTTCACGGAACCTTTATAGCACTCCGACATTTTCTGAGCAAGAGGAAGTCAAAATTTAT 1577
Db ||:|||||
QY 612 AACTTCACGGAACCTTTATAGCACTCCGACATTTTCTGAGCAAGAGGAAGTCAAAATTTAT 553
Db ||:|||||
QY 1578 TTAACACCTTAAGCCTTTTGTAGACTCTTTTCTATATATATGCTTAGGCTCACCATAGCGA 1637
Db ||:|||||
QY 552 TTAACACCTTAAGCCTTTTGTAGACTCTTTTCTATATATATGCTTAGGCTCACCATAGCGA 493
Db ||:|||||
QY 1638 ATTCTCCAGTGTAAACACTTTTCTGTTTTCACATTTGACTTTATGGGTTTGGGATTT 1697
Db ||:|||||
QY 492 ATTCTCCAGTGTAAACACTTTTCTGTTTTCACATTTGACTTTATGGGTTTGGGATTT 433
Db ||:|||||
QY 1698 TCTTGATGTTCTTATATATCCCHATATATATATATATATATGCAAAATTTTACATGTCAG 1757
Db ||:|||||
QY 432 TCTTGATGTTCTTATATATCCCHATATATATATATATATATGCAAAATTTTACATGTCAG 373
Db ||:|||||
QY 1758 CTACATGTTGGTAAGACACAGGCAAGTATTACTGTAAGTAAAGTATTTTAAAGTAA 1817
Db ||:|||||
QY 372 CTACATGTTGGTAAGACACAGGCAAGTATTACTGTAAGTAAAGTATTTTAAAGTAA 313
Db ||:|||||
QY 1818 ATATATTTTACGTGCTTTTGGCTTTTATTCAGAGTCTACATTTTATAGATTTCTACAT 1877
Db ||:|||||
QY 312 ATATATTTTACGTGCTTTTGGCTTTTATTCAGAGTCTACATTTTATAGATTTCTACAT 253
Db ||:|||||
QY 1878 CAGATGTTGTCACCTTATTTCCATTGGGATTCCTATTGTAAGTGTGTATGTCGCTGTTGG 1937
Db ||:|||||
QY 252 CAGATGTTGTCACCTTATTTCCATTGGGATTCCTATTGTAAGTGTGTATGTCGCTGTTGG 193
Db ||:|||||
QY 1938 AAAAGTGATTCATACCTTAGTTTTTTTCTCATCTGTTATCATACTTTTAAACAGCAAC 1997
Db ||:|||||
QY 192 AAAAGTGATTCATACCTTAGTTTTTTTCTCATCTGTTATCATACTTTTAAACAGCAAC 133
Db ||:|||||
QY 1998 CAATAACGGATGTAAAGTGTAAAGGCACAGGTTTACTCATGATGCTTCTGCAGACTGT 2057
Db ||:|||||
QY 132 CAATAACGGATGTAAAGTGTAAAGGCACAGGTTTACTCATGATGCTTCTGCAGACTGT 73
Db ||:|||||
QY 2058 GGGCTACACCACATATGTTATTGGGAAATATAGGTATTTTAGTACAGTACATACACT 2112
Db ||:|||||
```

```
|||||:|||||
72 GGGCTACACBACATATGTTATTGGAAATATAGGTATTTTAGTACAGTACATACT 18
|||||:|||||

BI916234      889 bp      mRNA      linear      EST 16-OCT-2001
603178134F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242305 5',
mRNA sequence.
BI916234
BI916234.1 GI:16180195
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 889)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11611 row: c column: 10
High quality sequence stop: 865.
Location/Qualifiers
1. 889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5242305"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."

BASE COUNT      214 a      264 c      226 g      185 t
ORIGIN

Query Match      31.3%; Score 795.8; DB 13; Length 889;
Best Local Similarity 98.5%; Pred. No. 7.8e-152;
Matches 846; Conservative 0; Mismatches 7; Indels 6; Gaps 4;

QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCCCTTCACGCCGCTCGGAGCGG 60
Db ||:|||||
QY 29 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCCCTTCACGCCGCTCGGAGCGG 88
Db ||:|||||
QY 61 ATAATCGGTTGAGCAGGCACCCAGCCGCGCAGACTCGGCTGGATCTGCGCAGACAGCGG 120
Db ||:|||||
QY 89 ATAATCGGTTGAGCAGGCACCCAGCCGCGCAGACTCGGCTGGATCTGCGCAGACAGCGG 148
Db ||:|||||
QY 121 GATTGCGTGGCCCGGGAGGAGCCCGGCGAGCGGCTGGGATCTCAGCGCGCGCGGTT 180
Db ||:|||||
QY 149 GATTGCGTGGCCCGGGAGGAGCCCGGCGCAGCGGCTGGGATCTCAGCGCGCGCGGTT 208
Db ||:|||||
QY 181 TGTCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCAGTTGGC 240
Db ||:|||||
QY 209 TGTCTGTTGTTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCAGTTGGC 268
Db ||:|||||
QY 241 CGTAACCTCAGGAAGTTCTCTTTTGACCCCATCTCATTTTCGAGCCACTTCTGAAGCCAC 300
Db ||:|||||
```



Db 269 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 328

QY 301 TTGAGAAAAATGATGTGACAGTTCTCTATCAAAAAGGATTCAGAAACATATACCATCTGTG 360

Db 329 TTGAGAAAAATGATGTGACAGTTCTCTATCAAAAAGGATTCAGAAACATATACCATCTGTG 388

QY 361 AAGAAAGTGGCCCTTTCTCCGCTTGCAAAATAGACATCTCAAAATTCCAAATGCCAGC 420

Db 389 AAGAAAGTGGCCCTTTCTCCGCTTGCAAAATAGACATCTCAAAATTCCAAATGCCAGC 448

QY 421 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAAGAAATTAACCTGAG 480

Db 449 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAAGAAATTAACCTGAG 508

QY 481 GGACATTTCTCTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 540

Db 509 GGACATTTCTCTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 568

QY 541 CATTTGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACATA 600

Db 569 CATTTGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACATA 628

QY 601 CGAGCTTTTACCTGGAAACCCAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 660

Db 629 CGAGCTTTTACCTGGAAACCCAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 688

QY 661 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAACGGCCTCCCGGTGCT 720

Db 689 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAACGGCCTCCCGGTGCT 748

QY 721 CAAAAATGCCATCT-CCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTAT 779

Db 749 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTAT 808

QY 780 TGTACACAGTGACATTTAA--TTCCAAACAGGAGTCTTC--GGGCCAGCAAAGCTGCC 835

Db 809 TGTACACAGTGACATTTAAATTTCCAAACAGGAGTCTTC--GGGCCAGCAAAGCTGCC 868

QY 836 AGG-CTTAGCTGCGAGGCC 853

Db 869 AGGCTTAGCTGCGAGGCC 887

RESULT 11

BI760499

LOCUS 603045037F1 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5185256 5',

DEFINITION mRNA sequence.

ACCESSION BI760499

VERSION BI760499.1 GI:15752077

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 870)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L1AM11462 row: j column: 09  
High quality sequence stop: 865.

FEATURES

source Location/Qualifiers

1..870

/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="IMAGE:5185256"  
/clone\_lib="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."  
BASE COUNT 212 a 247 c 228 g 182 t 1 others  
ORIGIN

Query Match 31.0%; Score 789.8; DB 13; Length 870;  
Best Local Similarity 98.2%; Pred. No. 1.3e-150;  
Matches 819; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTACGCCGTCTCGGAGCGG 60

Db 24 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTACGCCGTCTCGGAGCGG 83

QY 61 ATAATGCGGTGAGCAGGCACACCGCGGCAGACTCGGCTGGATCTCGCACAGCGGAGG 120

Db 84 ATAATGCGGTGAGCAGGCACACCGCGGCAGACTCGGCTGGATCTCGCACAGCGGAGG 143

QY 121 GATTGCGTGCGCCCGGGAGGCGCCCGGGGAGGCTGGGATCTTCAAGAGCCACTTCTGAAGCCAC 300

Db 144 GATTGCGTGCGCCCGGGAGGCGCCCGGGGAGGCTGGGATCTTCAAGAGCCACTTCTGAAGCCAC 323

QY 181 TGTCTGCTGTGTTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCCTCACTTGGC 240

Db 204 TGTCTGCTGTGTTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCCTCACTTGGC 263

QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGAGAGCCACTTCTGAAGCCAC 300

Db 264 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGAGAGCCACTTCTGAAGCCAC 323

QY 301 TTGAGAAAAATGATGTGACAGTCTCTATCAAAAAGGATTCAGAAACATATACCATCTGTG 360

Db 324 TTGAGAAAAATGATGTGACAGTCTCTATCAAAAAGGATTCAGAAACATATACCATCTGTG 383

QY 361 AAGAAAGTGGCCCTTTCTCCGCTTGCAAAATAGACATCTCAAAATTCCAAATGCCAGC 420

Db 384 AAGAAAGTGGCCCTTTCTCCGCTTGCAAAATAGACATCTCAAAATTCCAAATGCCAGC 443

QY 421 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAAGAAATTAACCTGAG 480

Db 444 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAAGAAATTAACCTGAG 503

QY 481 GGACATTTCTCTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 540

Db 504 GGACATTTCTCTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 563

QY 541 CATTTGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACATA 600

Db 564 CATTTGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACATA 623

QY 601 CGAGCTTTTACCTGGAACCCAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 660

Db 624 CGAGCTTTTACCTGGAACCCAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCATAATGA 683

QY 661 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAACGGCCTCCCGGTGCT 720

Db 684 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAACGGCCTCCCGGTGCT 743

QY 721 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTAT 780

Db 744 CAAAAATGCCATCTCCCTCCCGACATTTGGAGGAT-CCAAAGCTCTCATGTTGCCCTTA-T 801

QY 781 GTCACCAAGTACATTTAATTCACAAACAGGAGTCTTCGGGGCCAGCAAGAGCTGCC 834  
|||||  
Db 802 GTCACCAAGTACATTTAATTCACAAACAGGAGTCTTCGGGGCCAGCAAGAGCTGCC 855

RESULT 12

BI918195 827 bp mRNA linear EST 16-OCT-2001  
LOCUS 603183146F1 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5247130 5',  
DEFINITION mRNA sequence.

ACCESSION BI918195  
VERSION BI918195.1 GI:16181952  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 827)

REFERENCE

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11623 row: 1 column: 11

High quality sequence stop: 766.

FEATURES

source

Location/Qualifiers

1..827

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5247130"

/clone\_lib="NIH\_MGC\_121"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3  
fetal brains, female age 20 weeks, female age 24 weeks,  
and male age 26 weeks. Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range  
0.7-3.5 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 017. Note:  
this is a NIH\_MGC Library."

204 a 237 c 213 g 173 t

BASE COUNT

ORIGIN

Query Match 30.8%; Score 783.6; DB 13; Length 827;

Best Local Similarity 99.2%; Pred. No. 2.4e-149;

Matches 819; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGCCCTGGGGCGCCTTCACGCCGCTCTCGGAGCGG 60

|||||

Db 4 CGGGGCTAGCCCGGAGACCCGGCCACCGCCCTGGGGCGCCTTCACGCCGCTCTCGGAGCGG 63

|||||

QY 61 ATAATGCGGTGAGCAGGACACCGCCGCGGACACTCGGCTGGATCTCGGCACAGCGGCAGG 120

|||||

Db 64 ATAATGCGGTGAGCAGGACACCGCCGCGGACACTCGGCTGGATCTCGGCACAGCGGCAGG 123

|||||

QY 121 GATTGCGTGGCCCGCGGGAGCGCCGGGCGGACCGGCTGGGATCCTCAGCGCGCGCGGTT 180

|||||

Db 124 GATTGCGTGGCCCGCGGGAGCGCCGGGCGGACCGGCTGGGATCCTCAGCGCGCGCGGTT 183

|||||

QY 181 TGTCTGTTGTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCACTTGGC 240

|||||

Db 184 TGTCTGTTGTGGTCAAGACTGGATGATGTAAGTGGCTCTCTAGGAAGCCTCACTTGGC 243

QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGAAAGCCACTTCTGAAGCCAC 300  
|||||  
Db 244 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGAAAGCCACTTCTGAAGCCAC 303  
|||||  
QY 301 TTGAGAAAAATGATGTGACAGTTCCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 360  
|||||  
Db 304 TTGAGAAAAATGATGTGACAGTTCCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 363  
|||||  
QY 361 AAGAAAGTGGCCCTTTCTCCCGTTTGAATAATAGACATTCTCAAAATCCAAAATGCCAGC 420  
|||||  
Db 364 AAGAAAGTGGCCCTTTCTCCCGTTTGAATAATAGACATTCTCAAAATCCAAAATGCCAGC 423  
|||||  
QY 421 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAATTTAAACTGAG 480  
|||||  
Db 424 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAATTTAAACTGAG 483  
|||||  
QY 481 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 540  
|||||  
Db 484 GGACATTCTGTCTCTGATATGATCAGTCCCGCTTGGAGACTTTCGCCACACCATCCA 543  
|||||  
QY 541 CATTTGGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACTA 600  
|||||  
Db 544 CATTTGGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTTTTCAAGGGAACTA 603  
|||||  
QY 601 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCTTGGGCATAATGA 660  
|||||  
Db 604 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTTCCTTGGGCATAATGA 663  
|||||  
QY 661 GTTCTTCCGGGCCAACAGCAGCCTCGGACTCTGTGTTTCACAGAAACGCCCTCCCGGTGCT 720  
|||||  
Db 664 GTTCTTCCGGGCCAACAGCAGCCTCGGACTCTGTGTTTCACAGAAACGCCCTCCCGGTGCT 723  
|||||  
QY 721 CAAAATGCCATCTCCCTCCGACCATTTGAGGATCCCAAGCTCTCATGTTGCCCTTATT 780  
|||||  
Db 724 CAAAATGCCATCTCCCTCCGACCATTTGAGGATCCCAAGCTCTCATGTTGCCCTTATT 782  
|||||  
QY 781 GTCACCAAGTACATTTAATTCACAAACAGGAGTCTTCGGGGCCAGCA 826  
|||||  
Db 783 GTCA-CAGTGACATTTAA-TCCAAACAGGAGTCTTCGGGGCCAGCA 826

RESULT 13

BQ680912

LOCUS

DEFINITION

BQ680912 869 bp mRNA linear EST 15-JUL-2002

AGENCOURT\_8197216 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6261804

5', mRNA sequence.

ACCESSION

BQ680912

VERSION

BQ680912.1 GI:21793591

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 869)

REFERENCE

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: DCTD/DFP

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2423 row: b column: 13

High quality sequence stop: 719.

Location/Qualifiers

1..869

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6261804"

/clone\_lib="NIH\_MGC\_112"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."  
BASE COUNT 255 a 154 c 166 g 275 t 19 others  
ORIGIN

Query Match	30.8%;	Score 783;	DB 14;	Length 869;
Best Local Similarity	96.5%;	Pred. No. 3.2e-149;		
Matches	802;	Conservative 0;	Mismatches 26;	Indels 3;
Gaps	1;			
QY	1042	ATGCGAGCTCATCAAGGGAAGACTAAGTCAGAGGAGTCCCTCTCTGACCTTACAGGTTT	1101	
Db	1	ATGCGAGCTCATCAAGGGAAGACTAAGTCAGAGGAGTCCCTCTCTGACCTTACAGGTTT	60	
QY	1102	CCTCTCTCCCTGCAGCTTGATCTTGGGCCCTCCTCTCTGATGAGGCTGCTGATGTAAT	1161	
Db	61	CCTCTCTCCCTGCAGCTTGATCTTGGGCCCTCCTCTCTGATGAGGCTGCTGATGTAAT	120	
QY	1162	GGATAAAATAAGTAACAAGATGCCAACTTTTCCCTTGGGGTAAAGGTACAAAACA	1221	
Db	121	GGATAAAATAAGTAACAAGATGCCAACTTTTCCCTTGGGGTAAAGGTACAAAACA	180	
QY	1222	AACTAACACACAGTTGAAGAGAGGGCTTCCGAGCTGTATTTGCGAGTTTGTGGGTT	1281	
Db	181	AACTAACACACAGTTGAAGAGAGGGCTTCCGAGCTGTATTTGCGAGTTTGTGGGTT	240	
QY	1282	TTCTAAATAAATAATCTTACAAAGTATTTTACCTGTATGCTCTGCTTGCATAAACA	1341	
Db	241	TTCTAAATAAATAATCTTACAAAGTATTTTACCTGTATGCTCTGCTTGCATAAACA	300	
QY	1342	ATTAGAAAAAACAACAAGCAAAACCTATCTTGGCAAAAAAGGAAGTGCAGAGC	1401	
Db	301	ATTAGAAAAAACAACAAGCAAAACCTATCTTGGCAAAAAAGGAAGTGCAGAGC	360	
QY	1402	CCATTTTCAGGAGGCATTTGGTGATGTCGGCTCACATATTTGTCAGACACACAAGAAA	1461	
Db	361	CCATTTTCAGGAGGCATTTGGTGATGTCGGCTCACATATTTGTCAGACACACAAGAAA	420	
QY	1462	TCTGGCTTGGCCAGGATTTGGCAGTATGAAGGGCTGAGCGAGTCACATTAAGGAAT	1521	
Db	421	TCTGGCTTGGCCAGGATTTGGCAGTATGAAGGGCTGAGCGAGTCACATTAAGGAAT	480	
QY	1522	TCACGGAACCTTTATAGCACTCCGACATTTCTGAGCAAGAGGAACTTATTTTAA	1581	
Db	481	TCACGGAACCTTTATAGCACTCCGACATTTCTGAGCAAGAGGAACTTATTTTAA	540	
QY	1582	CACCTAAGCCTTTTGTAGACTCTTTTCTATATATTTAGGCTCACCATAGCAATTC	1641	
Db	541	CACCTAAGCCTTTTGTAGACTCTTTTCTATATATTTAGGCTCACCATAGCAATTC	600	
QY	1642	TCCAGTGTAAAACCTTTCTGTTTTCACATTTGAACCTTTATGGGTTTGGGATTTCTT	1701	
Db	601	TCCAGTGTAAAACCTTTCTGTTTTCACATTTGAACCTTTATGGGTTTGGGATTTCTT	660	
QY	1702	GTAGTTCTTATATATCCCTATATATATATATATATTTGCAAAATTTGACTGTCAGTAC	1761	
Db	661	GTAGTTCTTATATATCCCTATATATATATATATTTGCAAAATTTGACTGTCAGTAC	720	
QY	1762	ATGTTGGTAAGACACAGGCAAGTATTACTGTAAGTATTTTAAAGTTAAATAT	1821	
Db	721	ATGTTGGTAAGACACAGGCAAGTATTACTGTAAGTATTTTAAAGTTAAATAT	780	
QY	1822	ATTTT---ACGTGCTTTGGCTTTTATTTGCGAGTCTACATTTTATAGA	1869	
Db	781	ATTTTACGTGCCCTTTTGGCTTTTAAANNNNNNNNNNAANNNA	831	

RESULT 14  
BI915037  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
1. .786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5241534"  
/clone\_lib="NIH\_MGC\_121"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3  
fetal brains, female age 20 weeks, female age 24 weeks,  
and male age 26 weeks. Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range  
0.7-3.5 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 017. Note:  
this is a NIH\_MGC Library."  
BASE COUNT 195 a 225 c 192 g 174 t  
ORIGIN

Query Match	30.4%;	Score 774;	DB 13;	Length 786;
Best Local Similarity	99.9%;	Pred. No. 2.2e-147;		
Matches	785;	Conservative 0;	Mismatches 0;	Indels 1;
Gaps	1;			
QY	19	CCGCCCACCGGCTGGGGCCCTTACGCCGCTCTCGGAGCGGATAATCGGTGAGCAGGC	78	
Db	1	CCGCCCACCGGCTGGGGCCCTTACGCCGCTCTCGGAGCGGATAATCGGTGAGCAGGC	60	
QY	79	ACCACGCCGGCAGACTCGGTGGATCTGCGCACAGCGCAGGGATTGCGTGGCCCGCGG	138	
Db	61	ACCACGCCGGCAGACTCGGTGGATCTGCGCACAGCGCAGGGATTGCGTGGCCCGCGG	120	
QY	139	GAGGCCCGGGCAGCGGCTGGGATCTCAGCGCGCGGGTTGTCTGTTGTGTTCAA	198	
Db	121	GAGGCCCGGGCAGCGGCTGGGATCTCAGCGCGCGGGTTGTCTGTTGTGTTCAA	180	
QY	199	GACTGGATGATGTAAGTGGCTCTTAGGAAGCCCTACCTGGCCGTAACCTCAGGAAGTT	258	
Db	181	GACTGGATGATGTAAGTGGCTCTTAGGAAGCCCTACCTGGCCGTAACCTCAGGAAGTT	240	
QY	259	CTCTTTGACCCCATCTCATTTTGAAGCCACTTCTGAAGCCACTTGAGAAAAATGATGTA	318	
Db	241	CTCTTTGACCCCATCTCATTTTGAAGCCACTTCTGAAGCCACTTGAGAAAAATGATGTA	300	



```
QY 319 CAGTTCCTATCAAAAGGATTCAGAAACATATACCATCTGTGAAGAAAGTGCCCTTTCT 378
Db |||||||
QY 301 CAGTTCCTATCAAAAGGATTCAGAAACATATACCATCTGTGAAGAAAGTGCCCTTTCT 360
Db |||||||
QY 379 CCGGCTGCAAAATAGACATTTCTCAAAATTCCTCAAAATGCGGAGCCCAAGCCCAATTTACCT 438
Db |||||||
QY 361 CCGGCTGCAAAATAGACATTTCTCAAAATTCCTCAAAATGCGGAGCCCAAGCCCAATTTACCT 420
Db |||||||
QY 439 GAAAGCAGCCCAATACAGAAAGGAAAGAAATTTAAACTGAGGACATTTCTGTCTCTCTGA 498
Db |||||||
QY 421 GAAAGCAGCCCAATACAGAAAGGAAAGAAATTTAAACTGAGGACATTTCTGTCTCTCTGA 480
Db |||||||
QY 499 TATGATCAGTCCCCCGCTTGGAGACTTTTCGCCACACCATCCACATTTGCCAAAGAGGCCA 558
Db |||||||
QY 481 TATGATCAGTCCCCCGCTTGGAGACTTTTCGCCACACCATCCACATTTGCCAAAGAGGCCA 540
Db |||||||
QY 559 GCACGATGTCTTTGGAGATATTTCCCTTTCTTCAAGGGAACCTACGAGCTTTTACCTGGAAA 618
Db |||||||
QY 541 GCACGATGTCTTTGGAGATATTTCCCTTTCTTCAAGGGAACCTACGAGCTTTTACCTGGAAA 600
Db |||||||
QY 619 CCAGGAGAAAGCACACCTGGGCCAGTTCCTCTGGGCATATAGATTCTTCCGGGCCAACAG 678
Db |||||||
QY 601 CCAGGAGAAAGCACACCTGGGCCAGTTCCTCTGGGCATATAGATTCTTCCGGGCCAACAG 660
Db |||||||
QY 679 CACCTCGGACTCTGTGTTACAGAAACGCCCTCCCGGTGCTCAAAATGCCATCTCCCT 738
Db |||||||
QY 661 CACCTCGGACTCTGTGTTACAGAAACGCCCTCCCGGTGCTCAAAATGCCATCTCCCT 720
Db |||||||
QY 739 CCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTTATTGTCACCATGACATTTA 797
Db |||||||
QY 721 CCCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTTATTGTCACCATGACATTTA 780
Db |||||||
QY 798 ATTCCA 803
Db |||||||
QY 781 ATTCCA 786
Db |||||||
```

```
RESULT 15
BQ878634
LOCUS BQ878634
DEFINITION AGENCOURT_8113255 Lupski_dorsal_root_ganglion Homo sapiens cDNA
clone IMAGE:6178861 5', mRNA sequence.
ACCESSION BQ878634
VERSION BQ878634.1 GI:22270642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 959)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13559 row: b column: 14
High quality sequence start: 37
High quality sequence stop: 613.
Location/Qualifiers
1.959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6178861"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
```

```
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 261 a 264 c 221 g 213 t
ORIGIN
Query Match 30.3%; Score 770.2; DB 14; Length 959;
Best Local Similarity 95.4%; Pred. No. 1.2e-146;
Matches 869; Conservative 0; Mismatches 33; Indels 9; Gaps 7;
QY 307 AAATGATGTGACAGTTCTCTATCAAAAGGATTCAGAAACATATACCATCTGTGAAGAAA 366
Db |||||||
QY 45 AATATCATGTGACAGTTCTCTATC-AAAGGATTCAGAAACATATACCATCTGTGAAGAAA 103
Db |||||||
QY 367 GTGGCCCTTTCTCCCGCTTGCATAAATAGACATTTCTCAAAATTCCTCAAAATGCCAGCAAGAC 426
Db |||||||
QY 104 GTGGCCCTTTCTCCCGCTTGCATAAATAGACATTTCTCAAAATTCCTCAAAATGCCAGCAAGAC 163
Db |||||||
QY 427 CCCAATTTACCTGAAAGCAGCCCAATACAGAAAGGAAAGAAATTTAAACTGAGGGACAT 486
Db |||||||
QY 164 CCCAATTTACCTGAAAGCAGCCCAATACAGAAAGGAAAGAAATTTAAACTGAGGGACAT 223
Db |||||||
QY 487 TCTGTCTCTGATATGATCAGTCCCGCTTGGAGACTTTTCGCCACACCATCCACATTTGG 546
Db |||||||
QY 224 TCTGTCTCTGATATGATCAGTCCCGCTTGGAGACTTTTCGCCACACCATCCACATTTGG 283
Db |||||||
QY 547 CAAAGAGGGCCAGCAGATGTTCTTTGGAGATATTTCTCTTCTTCAAGGGAACACTACGAGCT 606
Db |||||||
QY 284 CAAAGAGGGCCAGCAGATGTTCTTTGGAGATATTTCTCTTCTTCAAGGGAACACTACGAGCT 343
Db |||||||
QY 607 TTTACCTGGAAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCAATAATGAGTTCTT 666
Db |||||||
QY 344 TTTACCTGGAAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTGGGCAATAATGAGTTCTT 403
Db |||||||
QY 667 CCGGGCCCAACAGCAGCTCGGACTCTGTGTTACAGAAACGCCCTCCCGGTGCTCAAAA 726
Db |||||||
QY 404 CCGGGCCCAACAGCAGCTCGGACTCTGTGTTACAGAAACGCCCTCCCGGTGCTCAAAA 463
Db |||||||
QY 727 TGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATTGTACC 786
Db |||||||
QY 464 TGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATTGTACC 523
Db |||||||
QY 787 AGTGACATTTAATTCCTCAACAGGAGTCTTCCGGGCCAGCAAAAGCTGCCAGGCTTAGCTG 846
Db |||||||
QY 524 AGTGACATTTAATTCCTCAACAGGAGTCTTCCGGGCCAGCAAAAGCTGCCAGGCTTAGCTG 583
Db |||||||
QY 847 CGAGCCCGTCTATGGAGGAAAGGCTCAGGAGAAAGCAGTCTGTGGAGAAATGGGACAGT 906
Db |||||||
QY 584 CGAGCCCGTCTATGGAGGAAAGGCTCAGGAGAAAGCAGTCTGTGGAGAAATGGGACAGT 643
Db |||||||
QY 907 CCACCA-GGAGACACCTCGTGGGCTCCAGCGGTTCTGCAATCTCAGTCCAG-CCAAGGC 964
Db |||||||
QY 644 CCACCAAGGAGAGACACCTCGTGGGCTCCAGCGGTTCTGCAATCTCAGTCCAGCCCGGSC 703
Db |||||||
QY 965 AGAGACAGCCACTCTCCAGCTGTCCGAACAGTACCCCGACTGCCAGCCGAGGACATG 1024
Db |||||||
QY 704 AGAGACAGCCACTCTCCAGCTGTCCGAACAGTACCCCGACTGCCAGCCCGAGGACATG 763
Db |||||||
QY 1025 TTTGACCATCCACCCCATCGGAGCTCATCAAGGG-AAAGACTAAGTCAGAGGAGTCCCT 1083
Db |||||||
QY 764 TTTGACCATCCACCCCATCGGAGCTCATCAAGGGAAAGACTAAGTCAGAGGAGTCCCT 823
Db |||||||
QY 1084 CTCT-GACCTTACAGGTTCCCTCTCTCCCTGCAGCTTG--ATCTTGGGCGCCTCACTTT 1140
Db |||||||
```

Db	824	CTCTGGACTTTACGGGTTCCCTCCTCCTCCTGGAGCTTGGATCCTGGGCCCTTCACTTTT	883
QY	1141	GGA--TGAGGTGCTGAATGTAATGGATAAAAAATAAGTAACAAGATGCCAACTTTTTTCCT	1198
Db	884	GGAATTGAGGGCTTAAATGGAATGGATAAAAAATAAGTAACCAGATGCCACTTTTTTTCC	943
QY	1199	TTGGGGTAAAA	1209
Db	944	TTTGGGGGTAA	954

Search completed: July 25, 2003, 19:53:57  
Job time : 3785.55 secs





Tue Jul 29 09:34:17 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 15:39:40 ; Search time 563.197 seconds  
(without alignments)  
9326.039 Million cell updates/sec

Title: US-09-806-276A-4  
Perfect score: 2546  
Sequence: 1 cggggctagccggagacc.....cagctanaacgagggcagta 2546

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues  
Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq1:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq3:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1372	53.9	1832	15	US-10-153-668-16
2	441.4	17.3	486	11	US-09-796-692-7465
3	441.4	17.3	486	15	US-10-040-862-7465
4	374.6	14.7	445	12	US-09-918-995-5286
5	301	11.8	301	11	US-09-796-692-7516
6	301	11.8	301	15	US-10-040-862-7516
7	90.4	3.6	177	11	US-09-783-590-9071
8	85	3.3	413	10	US-09-864-761-26628
9	85	3.3	467	10	US-09-864-761-9983
10	82.4	3.2	169	11	US-09-783-590-6760
11	81.6	3.2	145	11	US-09-783-590-5334
12	62	2.4	11812	15	US-10-239-676-210
13	56.8	2.2	9539	15	US-10-239-676-52
14	53.6	2.1	416	11	US-09-960-352-4584
15	52.4	2.1	516	11	US-09-960-352-5785
16	52.2	2.1	446	11	US-09-960-352-3400

17	52.2	2.1	529	11	US-09-983-965-2109	Sequence 2109, Ap
18	51.6	2.0	9539	15	US-10-239-676-51	Sequence 51, Appl
19	50.6	2.0	525	15	US-10-198-846-1483	Sequence 1483, Ap
20	49.8	2.0	344	11	US-09-960-352-1036	Sequence 1036, Ap
21	49.4	1.9	393	11	US-09-960-352-4582	Sequence 4582, Ap
22	49.2	1.9	11836	15	US-10-239-676-102	Sequence 102, Appl
23	48.8	1.9	13606	15	US-10-239-676-165	Sequence 165, Appl
24	48.4	1.9	8093	15	US-10-172-086-25	Sequence 25, Appl
25	48.2	1.9	411	11	US-09-960-352-14521	Sequence 14521, A
26	48.2	1.9	480	11	US-09-960-352-5301	Sequence 5301, Ap
27	48	1.9	15732	15	US-10-239-676-96	Sequence 96, Appl
28	47.6	1.9	431	11	US-09-960-352-5558	Sequence 5558, Ap
29	47.4	1.9	10619	15	US-10-239-676-1	Sequence 1, Appl
30	47.2	1.9	8333	15	US-10-239-676-113	Sequence 113, Appl
31	47.2	1.9	15732	15	US-10-239-676-95	Sequence 95, Appl
32	47	1.8	6046	15	US-10-239-676-16	Sequence 16, Appl
33	46.4	1.8	11036	15	US-10-239-676-118	Sequence 118, Appl
34	46.2	1.8	5917	15	US-10-087-464-9	Sequence 9, Appl
35	46.2	1.8	12968	15	US-10-239-676-202	Sequence 202, Appl
36	46.2	1.8	14147	15	US-10-172-086-52	Sequence 52, Appl
37	46	1.8	5689	15	US-10-239-676-90	Sequence 90, Appl
38	46	1.8	6352	15	US-10-172-086-24	Sequence 24, Appl
39	46	1.8	11260	15	US-10-239-676-19	Sequence 19, Appl
40	45.8	1.8	6306	15	US-10-239-676-224	Sequence 224, Appl
41	45.8	1.8	6944	15	US-10-172-086-112	Sequence 112, Appl
42	45.6	1.8	7657	15	US-10-239-676-185	Sequence 185, Appl
43	45.6	1.8	513509	12	US-09-754-853A-4	Sequence 4, Appl
44	45	1.8	11036	15	US-10-239-676-117	Sequence 117, Appl
45	44.8	1.8	7823	15	US-10-239-676-198	Sequence 198, Appl

ALIGNMENTS

RESULT 1  
US-10-153-668-16  
; Sequence 16, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO.16  
; LENGTH: 1832  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (508)..(837)  
US-10-153-668-16

Query Match 53.9%; Score 1372; DB 15; Length 1832;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	2254	GTAAAGAGATGATTCCACTACTCTCATTTTGATTACTTCCCTTATGGTTTGAGAGGCTAG	2313
Db	1081	GTAAAGAGATGATTCCACTACTCTCATTTTGATTACTTCCCTTATGGTTTGAGAGGCTAG	1140
QY	2314	AAACTGAGCCTCTCTACTTTTGGAAAAATGAACATGTGAGGTGAGATTTTCTTTT	2373
Db	1141	AAACTGAGCCTCTCTACTTTTGGAAAAATGAACATGTGAGGTGAGATTTTCTTTT	1200
QY	2374	TTTAAGTCAGCACTGATGCCACCCTCTCAGTGGTCATTTCTGAGCATCTTCTGACTTGA	2433
Db	1201	TTTAAGTCAGCACTGATGCCACCCTCTCAGTGGTCATTTCTGAGCATCTTCTGACTTGA	1260
QY	2434	ACACCTTCTACAGCAAACTCTTGCAAGTCCAGTTTCATCCCTGTAAAGCAAAATGCTCTTT	2493
Db	1261	ACACCTTCTACAGCAAACTCTTGCAAGTCCAGTTTCATCCCTGTAAAGCAAAATGCTCTTT	1320
QY	2494	CACGCAGAAAAGTGCCATATAGACGAGATAAAGGCAGCTAATACAGGGGCAGTA	2546
Db	1321	CACGCAGAAAAGTGCCATATAGACGAGATAAAGGCAGCTAATACAGGGGCAGTA	1373

RESULT 2  
US-09-796-692-7465/c  
; Sequence 7465, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
FILE RECEIVED 0017

```

; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7465
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (175)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (271)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (394)

```

Query Match

PRIOR FILING DATE: 2000-08-04  
 PRIOR APPLICATION NUMBER: US 60/223,378  
 PRIOR FILING DATE: 2000-08-07  
 PRIOR APPLICATION NUMBER: US 09/796,692  
 PRIOR FILING DATE: 2001-03-01  
 NUMBER OF SEQ ID NOS: 10467  
 SOFTWARE: FastSEQ for Windows version 3  
 SEQ ID NO 7465

PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSEQ for Windows version 3  
SEQ ID NO 7465

LENGTH: 486  
TYPE: DNA  
ORGANISM: Homo sapiens

NAME/KEY: unsure  
LOCATION: (175)  
OTHER INFORMATION: n=A, T, C or G

NAME/KEY: unsure  
LOCATION: (271)  
OTHER INFORMATION: p=A, T, C or G

NAME/KEY: unsure  
LOCATION: (394)  
OTHER INFORMATION: n/a. T. C. or G.

FEATURE:  
NAME/KEY: unsure  
LOCATION: (395)  
OTHER INFORMATION: n-a c or c

FEATURE:  
NAME/KEY: unsure  
LOCATION: (415)

FEATURE:  
NAME/KEY: unsure  
LOCATION: (436)

FEATURE:  
NAME/KEY: unsure  
LOCATION: (446)

3-10-040-862-7465  
Query Match 17.3%; Score 44

Matches 445; Conservative 0; Mismatches 1000 CCCCAGCTGGCCAGCCGAGGACATGTTT

b  
459 CCCGGCAGGTCCANCCGAGGACANGTTT  
1060 AAAGACTAAGTCAGAGGAGTCCCTCTCTC

399 AAAGNNTAAGTCAGAGGAGTCCCCTCTC  
1120 TCATCTTGGGCCCTCACTTTTGGATGAA

b

339 TGATCTGGGCCCTCACTTTGGGATGA

1100 TCAATCCCAACCTTTTCCCTTCCCTCA

Gaps	0;
CAAGGG	1059
CAAGGG	400
GCAGCT	1119
GCAGCT	340
AGTAACA	1179
AGTAACA	280
TTGAAG	1239



Db 279 AGATGCCANCTTTTCTTCTTGGGTTAAAGGTACAAAACAACAACTAACACAGTTGAAG 220  
QY 1240 AGAAGGGCTTCCGAGCTGTATTGTCAGTTTGTGTTGGGTTTCTTCTAAATAATATCTT 1299  
Db 219 AGAAGGGCTTCCGAGCTGTATTGTCAGTTTGTGTTGGGTTTNTAAATAATATCTT 160  
QY 1300 ACAAGATTTTACCTGTTATGCCCTGTTTGCATAAACAATTTAGAAAAACAACA 1359  
Db 159 ACAAGATTTTACCTGTTATGCCCTGTTTGCATAAACAATTTAGAAAAACAACA 100  
QY 1360 AAGCAAAACCTATCTTGGCAAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCATT 1419  
Db 99 AAGCAAAACCTATCTTGGCAAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCATT 40  
QY 1420 GGTGATGTCGGCTCACATATTGTTGCAGACACACAA 1457  
Db 39 GGTGATGTCGGCTCACATATTGTTGCAGACACACAA 2

RESULT 4

US-09-918-995-5286  
; Sequence 5286, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseg, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5286  
; LENGTH: 445  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(445)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-5286

Query Match 14.7%; Score 374.6; DB 12; Length 445;  
Best Local Similarity 94.2%; Pred. No. 3.7e-87;  
Matches 389; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGCGCTTACGCCCGTCTCGGAGCGG 60  
Db 31 CGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGCGCTTACGCCCGTCTCGGAGCGG 90  
QY 61 ATAATGCGGTGACGAGGACACCGCCGCGAGACTCGGCTGGATCTGCGACAGCGGCGAG 120  
Db 91 ATAATGCGGTGACGAGGACACCGCCGCGAGACTCGGCTGGATCTGCGACAGCGGCGAG 150  
QY 121 GATTGCGTGGCGCCCGGGAGGCGCGGCGAGCGGCTGGATCTTCAAGCCGCGCGCGTT 180  
Db 151 GATTGCGTGGCGCCCGGGAGGCGCGGCGAGCGGCTGGATCTTCAAGCCGCGCGCGTT 210  
QY 181 TGTCCTGGTGTGGTCAAGACTGGATGATGTAACCTGGCTCTCTAGGAAGCCTCACTTGGC 240  
Db 211 TGTCCTGGTGTGGTCAAGACTGGATGATGTAACCTGGCTCTCTAGGAAGCCTCACTTGGC 270  
QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTCAAGCCACTTCTGAAGCCAC 300  
Db 271 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTCAAGCCACTTCTGAAGCCAC 330  
QY 301 TTGAGAAAAATGATGTGACAGTTCCTATCAAAAAGGATTTCAGAAACATATACCATCTGTG 360  
Db 331 TTGAGAAAAATGATGTGACAGTTCCTATCATAAAGGATTTCAGAAACATATACCATCTGTG 390

QY 361 AAGAAAGTGCCCTTTCTTCCGCTTGCAAAATAGACATTTCTCAAAATTCACAAA 413  
Db 391 AAGAAAGTGCCCTTTCTTCCGCTTTTGGGCAATAGAAATTTCTCATGGAATCTCTAAA 443

RESULT 5

US-09-796-692-7516/c  
; Sequence 7516, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; FILE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7516  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-796-692-7516

Query Match 11.8%; Score 301; DB 11; Length 301;  
Best Local Similarity 100.0%; Pred. No. 4.3e-68;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 709 CTCCCCGGTGCTCAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCAT 768  
Db 301 CTCCCCGGTGCTCAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCAT 242  
QY 769 GTTGCCCTTATTGTCACCAGTGACATTTAAATCCAAACAGGATCCTTCGGGCCAGCAA 828  
Db 241 GTTGCCCTTATTGTCACCAGTGACATTTAAATCCAAACAGGATCCTTCGGGCCAGCAA 182  
QY 829 GCTGCCCGAGGCTTAGCTCGGAGCCCGTCATGGAGGAAAAGCTCAGGAGAAAAGCAGTCT 888  
Db 181 GCTGCCCGAGGCTTAGCTCGGAGCCCGTCATGGAGGAAAAGCTCAGGAGAAAAGCAGTCT 122  
QY 889 GTTGAGAAATGGGACAGTCCACAGGAGACACCTCGTGGGCTCCAGCGGTTCTGCATC 948  
Db 121 GTTGAGAAATGGGACAGTCCACAGGAGACACCTCGTGGGCTCCAGCGGTTCTGCATC 62  
QY 949 TCAGTCCAGCCCAAGGACAGAGACAGCCACTCTCCAGCCTGTCCGAACAGTACCCCGACTG 1008  
Db 61 TCAGTCCAGCCCAAGGACAGAGACAGCCACTCTCCAGCCTGTCCGAACAGTACCCCGACTG 2

Db 121 GTTGGAGATGGGACAGTCCACCAGGGAGACACCTCGTGGGGCTCCAGCGGTTCTGCATC 62  
QY 949 TCAGTCCAGCCAAAGCAGAGACAGCCACTCTCCAGCCTGTCCGAAACAGTACCCCGACTG 1008  
Db 61 TCAGTCCAGCCAAAGCAGAGACAGCCACTCTCCAGCCTGTCCGAAACAGTACCCCGACTG 2  
QY 1009 G 1009  
Db 1 G 1

RESULT 7  
US-09-783-590-9071  
; Sequence 9071, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillion, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9071  
; LENGTH: 177  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (56)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (59)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (121)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (127)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (147)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-9071

Query Match 3.6%; Score 90.4; DB 11; Length 177;  
Best Local Similarity 89.9%; Pred. No. 2.4e-13;  
Matches 116; Conservative 0; Mismatches 11; Indels 2; Gaps 2;  
QY 78 CACCACGCCGCGGACAGTCCGCTGGATCTCGCACAGCGGCGGATTCGCTGCCCGCG 137  
Db 49 CTCTCCGNCGNCAGACTCGGCTGGATCTCGCACAGCGGACAGGATTCGCTGCCCGCG 108  
QY 138 GG-AGGCCCCGGGCGAGCGGCTGGG-ATCCTCAGCGCGCGCGGTTTGTCTGTGGT 195  
Db 109 GGAAGGCCCGGNCACGNCCTGGGAATCCTCAGCGCGCNCCTCGTTTGTCTGTGGT 168  
QY 196 CAAGACTGG 204  
Db 169 CAAGACTGG 177

RESULT 8  
US-09-864-761-26628  
; Sequence 26628, Application US/09864761

QY 1009 G 1009  
Db 1 G 1  
RESULT 6  
US-10-040-862-7516/c  
; Sequence 7516, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7516  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-7516

Query Match 11.8%; Score 301; DB 15; Length 301;  
Best Local Similarity 100.0%; Pred. No. 4.3e-68;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 709 CTCGCCGCTGCTCAAAATGCCATCTCCCTCCGACCATTTGAGGATCCCAAGCTCTCAT 768  
Db 301 CTCGCCGCTGCTCAAAATGCCATCTCCCTCCGACCATTTGAGGATCCCAAGCTCTCAT 242  
QY 769 GTTGCCCTTATGTACACAGTGACATTTAAATCCAAACAGGAGTCTTCGGGCCAGCAA 828  
Db 241 GTTGCCCTTATGTACACAGTGACATTTAAATCCAAACAGGAGTCTTCGGGCCAGCAA 182  
QY 829 GCTGCCAGGCTTAGCTGCGAGCCCGTCATGGAGGAAAAAGCTCAGGAGAAAGCATCT 888  
Db 181 GCTGCCAGGCTTAGCTGCGAGCCCGTCATGGAGGAAAAAGCTCAGGAGAAAGCATCT 122  
QY 889 GTTGGAGATGGGACAGTCCACCAGGGAGACACCTCGTGGGGCTCCAGCGGTTCTGCATC 948

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26628
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC00353.21
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.62
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 39
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EST_HUMAN HIT: BF058866.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q00587, EVALUE 3.00e-07
; OTHER INFORMATION: NT HIT: gill141318, EVALUE 0.00e+00
US-09-864-761-26628
```

```
Query Match 3.3%; Score 85; DB 10; Length 413;
Best Local Similarity 64.7%; Pred. No. 1.1e-11;
Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;
```

```
QY 409 CAAATGCCAGCCCAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAAGAA 468
```

```
Db 101 CACCATGTCCACCAAGGTGCCCATCTATCTGAAGCGTGGCAGTCGCAAG---GGCAAGAA 157
QY 469 ATTTAAACTGAGGGACATCTGTCTCTGATATGATAGTCCCGCTTGGAGACTTTCG 528
Db 158 GGAGAGCTTCGGGACCTGCTGCTCGGACATGATCAGCCCGCTGGGGGACTCCG 217
QY 529 CCACACCATCCACATTTGCAAGAGGGCCAGCAGCATGCTTTGGAGATATTTCTTCT 588
Db 218 CCACACCATTCATATTGGCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 277
QY 589 TCAAGGGAACTACGAGCTTTTACCTGGAAACCCAGGAGAAAG 629
Db 278 GCAGGGCAAGTTCACCTCCTGCCGGGACCATGGTGGAGG 318
```

## RESULT 9

```
US-09-864-761-9983
; Sequence 9983, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9983
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```



; NAME/KEY: misc feature

```
; NAME/KEY: misc feature
; LOCATION: (122)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (128)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5334

Query Match      3.2%; Score 81.6; DB 11; Length 145;
Best Local Similarity 89.1%; Pred. No. 4.1e-11;
Matches 106; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 966 GAGACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGACTGGCCAGCCGAGGACATGT 1025
Db 1 GGNAGAGCCACTCCTCCAGCCTNTCCGANCAGTACCCCGACTGGCCAGCCGAGGACATGT 60

QY 1026 TTGACCATCCACCCCATGCG-AGCTCATCAAGGG-AAAGACTAAAGTCAGAGGAGTCC 1082
Db 61 TTNACCATCCACCCCATGTGAAGNTCANCAAGGNNAAAGACTAAAGTCAGAGGAGTCC 119

RESULT 12
US-10-239-676-210
; Sequence 210, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 210
; LENGTH: 11812
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-210

Query Match      2.4%; Score 62; DB 15; Length 11812;
Best Local Similarity 47.2%; Pred. No. 0.00011;
Matches 286; Conservative 0; Mismatches 315; Indels 5; Gaps 3;

QY 1570 AAATTATTAAACACCTAAGCCCTTTTGTAGACTCTTTTCTATATATATGCTTAGGCTCAC 1629
Db 2370 AAATTATTTTAGTTTAAATAATATATAAAATAGGTTGTTATTGATTTTAGGTTATA 2429

QY 1630 CATAGCGAATCTCCAGTGTAAACCTTTTCTGTTTTCACATTTGAACCTTATGGGTTTT 1689
Db 2430 GTTTGTTGATCGTC--GTTTATTAAATTTTGAATAATATATATTTTAAATAGATTAA 2487

QY 1690 GGGGATTTCTGTAGTTCTTATATATCCCTATATATATATATATATATGCAAAATTTG 1749
Db 2488 GAATTTTCTTTGTTGTTTCTTTTGTGTTGTTTAAATAATAATAATGTTTATATG 2547

QY 1750 ACTGTCAGCTACATGTTGGTAAGACACAGGCAAGATTTACTGTAAGTAAAGTTATTTTA 1809
Db 2548 ATGATTTTTTAAAGTTTGAGTTTAAAAATTTTTTTTATTTTAAATATATGATAATTTT 2607

QY 1810 AAGTTAAATAATATTTTTTACGTGCCTTTGGCTTTTATTGCGAGAGTCTACATTTTATAGA 1869
```

```
Db 2608 ATTATGTATTTTAAAGAAAT-TTAAATTTAAATATGATAATTTTATTTTAAAGA 2666
QY 1870 TTCTACATCAGATGTTGTCACTTATTTCCATTTGGGATTCATTTGAAGCTGTGTATGTC 1929
Db 2667 ATTTTAATTTAAATATGATAATTTTATTTATTTATTTAAATTTTATTTATTTAG 2726
QY 1930 GTGTTTGGAAAAGTGTATTCATACCTTAGTTTCTTTCTTCTCATCTGTATCATCTTTTA 1989
Db 2727 ATTTTGTGTTTAAATTTTGTAAATATTTTGTGATGTTATTTTGTGTTTGTAGTTT 2786
QY 1990 ACAGCAACCAATAACGGGATTGTAAAGTGTAAAGGCACAGGTTACTCATGATGCTTCTGCA 2049
Db 2787 GGATATTTATTTTAGTTATTTTAAAGATTTTCTTTAGTTTATTTATTTAGTTTAA 2846
QY 2050 GAGACTGTGGCTACACCACATATGTTTATTTGGAAATATAGGTATTTTAGTACAGTACAT 2109
Db 2847 ATTAATTTTAAATATATATAATTTAGTGTAAATAT--TAATTTGGAAAAAAGTT 2904
QY 2110 ACTTGCAATTACATAGGTACTTCAAGCAACACAATAAAAAAGTAAATGATAAAGTGAAC 2169
Db 2905 ATTTATTTTGTAGGTTTGGTTAAGGATGATATAAAATGTTTAAATTTAGGGAATAAT 2964
QY 2170 CTTGTT 2175
Db 2965 GTTATT 2970

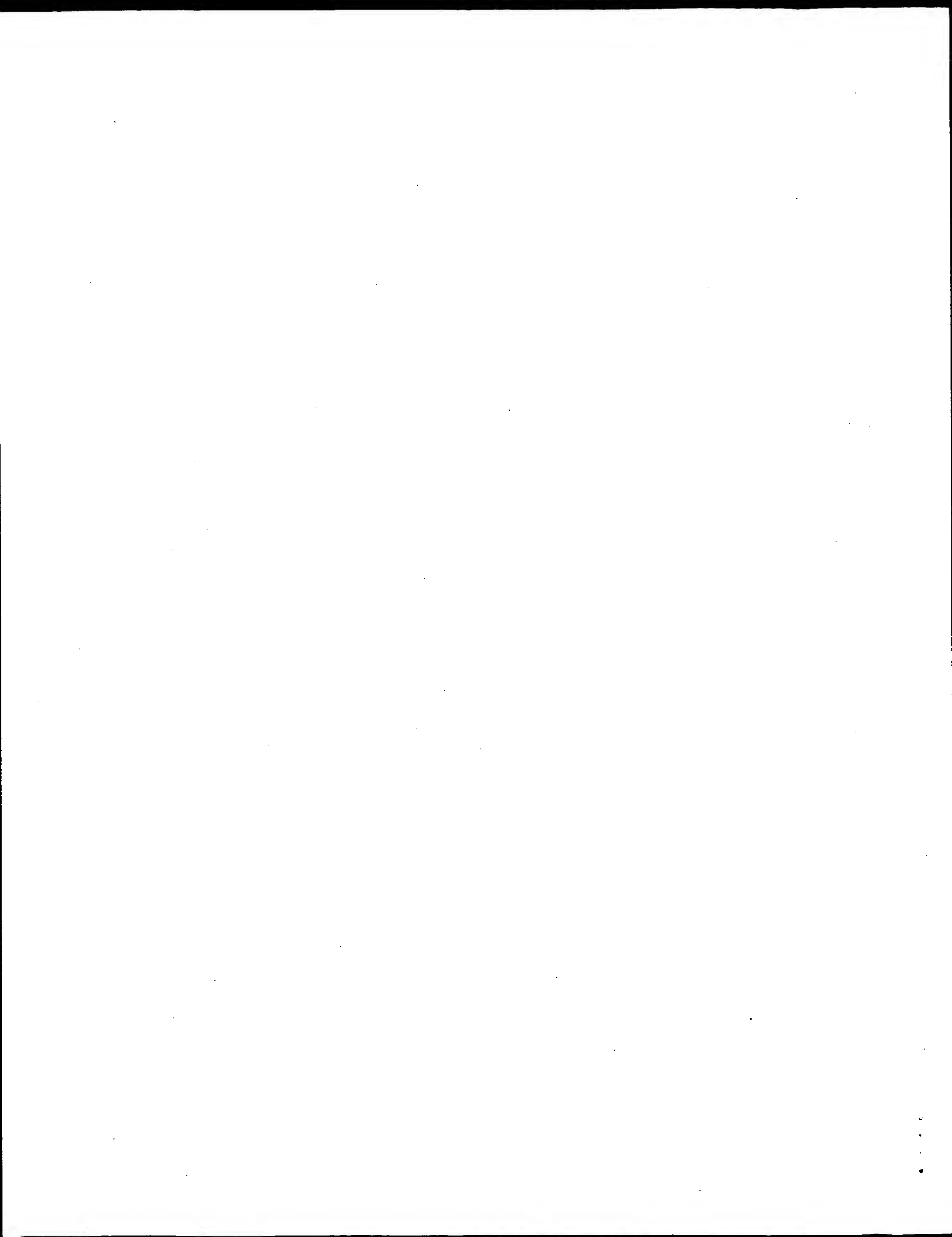
RESULT 13
US-10-239-676-52
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 52
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-52

Query Match      2.2%; Score 56.8; DB 15; Length 9539;
Best Local Similarity 46.1%; Pred. No. 0.0021;
Matches 190; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 1577 TTAAACACCTAAGCCCTTTTGTAGACTCTTTTCTATATATATGCTTAGGCTCACCATAGCG 1636
Db 21 TTTAGATATAGAAATTTTTTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTT 80
QY 1637 AATCTCCAGTGTAAACCTTTTCTGTTTTCACATTTGAACCTTTATGGGTTTGGGAT 1696
Db 81 TTTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 140
QY 1697 TTCTGTAGTTCTTATATATCCCTATATATATATATATATATATGCAAAATTTTGACTGTCA 1756
```







us-09-806-276a-4.rni

Tue Jul 29 09:34:16 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 25, 2003, 15:25:15 ; Search time 123.381 Seconds  
(without alignments)  
6328.365 Million cell updates/sec

Title: US-09-806-276a-4  
Perfect score: 2546  
Sequence: 1 cggggctagccggagagacc.....cagctanaacgagggcagta 2546

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52	2.0	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 2	45.4	1.8	7218	1 US-08-232-463-14	Sequence 14, Appl
C 3	44	1.7	6243	2 US-09-056-075-1	Sequence 1, Appl
C 4	42.8	1.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 5	41	1.6	8220	2 US-08-568-459A-11	Sequence 11, Appl
C 6	41	1.6	8220	2 US-08-487-826B-11	Sequence 11, Appl
C 7	41	1.6	8220	4 US-09-210-288-11	Sequence 11, Appl
C 8	40	1.6	2264	3 US-08-262-220-5	Sequence 5, Appl
C 9	40	1.6	2264	3 US-08-471-733-5	Sequence 5, Appl
C 10	40	1.6	2264	3 US-08-468-878-5	Sequence 5, Appl
C 11	40	1.6	2264	4 US-08-750-494-5	Sequence 5, Appl
C 12	39.2	1.5	660	1 US-07-991-867B-32	Sequence 32, Appl
C 13	39.2	1.5	660	1 US-08-107-755A-32	Sequence 32, Appl
C 14	39.2	1.5	660	2 US-08-544-332-32	Sequence 32, Appl
C 15	39.2	1.5	660	4 US-09-370-861A-32	Sequence 32, Appl
C 16	39.2	1.5	854	4 US-08-998-416-534	Sequence 534, App
C 17	39.2	1.5	1511	1 US-07-991-867B-8	Sequence 8, Appl
C 18	39.2	1.5	1511	1 US-08-107-755A-8	Sequence 8, Appl
C 19	39.2	1.5	1511	2 US-08-544-332-8	Sequence 8, Appl
C 20	39.2	1.5	1511	4 US-09-370-861A-8	Sequence 11, Appl
C 21	39.2	1.5	4810	3 US-08-852-629-11	Sequence 15, Appl
C 22	39.2	1.5	4838	3 US-08-852-629-15	Sequence 595, App
C 23	39	1.5	658	4 US-08-998-416-595	Sequence 16, Appl
C 24	38.8	1.5	473	1 US-08-764-100-16	Sequence 541, App
C 25	38.8	1.5	821	4 US-08-998-416-541	Sequence 315, App
C 26	38.8	1.5	1483	4 US-08-961-527-315	Sequence 14, Appl
C 27	38.8	1.5	4970	1 US-08-764-100-14	

28	38.8	1.5	4970	1 US-08-764-100-20	Sequence 20, Appl
29	38.6	1.5	6095	4 US-09-357-206A-18	Sequence 18, Appl
30	38.6	1.5	6325	4 US-09-357-206A-20	Sequence 20, Appl
31	38.6	1.5	6428	4 US-09-357-206A-22	Sequence 22, Appl
32	38.6	1.5	7400	1 US-08-261-663A-1	Sequence 1, Appl
33	38.6	1.5	7400	5 PCT-US95-07754A-1	Sequence 1, Appl
34	38.6	1.5	12286	4 US-09-357-206A-1	Sequence 3, Appl
35	38.6	1.5	38564	4 US-09-734-673-3	Sequence 651, App
36	38	1.5	20674	4 US-09-641-638-651	Sequence 3, Appl
37	37.4	1.5	3457	4 US-09-293-549-3	Sequence 2, Appl
38	37.2	1.5	5852	1 US-07-867-106-2	Sequence 2243, Ap
39	37.2	1.5	11091	4 US-09-134-001C-2243	Sequence 3, Appl
40	37.2	1.5	28001	4 US-09-819-993-3	Sequence 32, Appl
41	36.8	1.4	1850	3 US-08-617-860B-32	Sequence 4, Appl
42	36.8	1.4	2781	3 US-08-749-522-4	Sequence 4, Appl
43	36.8	1.4	4098	2 US-08-605-106-4	Sequence 3, Appl
44	36.6	1.4	1780	1 US-08-434-702-3	Sequence 3, Appl
45	36.6	1.4	1780	1 US-08-271-883-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-487-826B-13/c  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-8550

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 2.0%; Score 52; DB 2; Length 19124;  
Best Local Similarity 46.9%; Pred. No. 0.0011;

[illegible]

RESULT 3  
US-09-056-075-1/c  
; Sequence 1, Application US/09056075  
; Patent No. 5955368  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric A.  
; APPLICANT: Bradshaw, Marite  
; APPLICANT: Rood, Julian  
; TITLE OF INVENTION: Expression System for Clostridium  
; TITLE OF INVENTION: Species  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,075  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95238  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000

## RESULT 2



Tue Jul 29 09:34:16 2003

REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 1.7%; Score 42.8; DB 2; Length 19124;  
Best Local Similarity 51.2%; Pred. No. 0.34;  
Matches 124; Conservative 0; Mismatches 117; Indels 1; Gaps 1;  
QY 1732 CTATATTGCAAAATTTTGACTGTGCTACATGTTGGTAAGACACAGGCAAGTATTACT 1791  
Db 4282 CAACCTTGCATAATTTGTAATTAATAAATAATATATATAAAGAAATATATAAATAAT 4341  
QY 1792 GTAAGTAAAGTATTTTAAAGTTAAATATATATTTTACGTGCTTGGCTTTTATTGCA 1851  
Db 4342 ATAAATAAATAAGCATAAATGTCACATAAATTTTATTTTAAATTAATTTTATTTTA 4401  
QY 1852 GAG-TCTACATTTTATAGATTTCTACATCAGATGTTGTCACCTTATTTCCATTTGGGATTCCA 1910  
Db 4402 TTGTTCTAAATAATATATTGATTAAGAAATATTTTGTGCTCTAATAATAAATAAGATATTT 4461  
QY 1911 TTGTAAGCTGTGTATGTGCGTGTGTTGGAAAAGTGTATTCATCTAGTTTCTTCTTC 1970  
Db 4462 CTAATATTAATTTATATATAATATATAATTTTAAAGTATTTTAAAGTATTTTACTATT 4521  
QY 1971 AT 1972  
Db 4522 AT 4523

RESULT 5

US-08-568-459A-11  
Sequence 11, Application US/08568459A  
Patent No. 5849306  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655

TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from  
OTHER INFORMATION: plasmid RP4"  
US-09-056-075-1  
Query Match 1.7%; Score 44; DB 2; Length 6243;  
Best Local Similarity 47.2%; Pred. No. 0.087;  
Matches 134; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 1706 TTCTTATATATCCCTATATATATATATATATTTGCTGCTGCTACATGT 1765  
Db 1487 TTATAAGGCTCATTTTATATATCTTTCTTCAAGATATATATAAATAATTTT 1428  
QY 1766 TGGTAAGACACAGGCAAGTATTACTGTAAGTAAAGTTAAAGTTAAATATATTT 1825  
Db 1427 TTCAAACCTTAAATAAATAAATAATTTTATATATTTTATTTTATTTTATTTT 1368  
QY 1826 TTACGTGCTTTGGCTTTTATTTGTCAGAGTCTACATTTTATAGATTTCTACATCAGATGTT 1885  
Db 1367 TTATTTTATTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1308  
QY 1886 GTCACCTATTCCATTTGGGATTTCCATTTGTAAGCTGTGTATGCGTGTGTTGGAAGTGT 1945  
Db 1307 TTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1248  
QY 1946 ATTCATACCTAGTTTATTTTCTTCTCATCTGTTATCATCTTTTA 1989  
Db 1247 TTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1204

RESULT 4

US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655

```

; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-11

Query Match      1.6%; Score 41; DB 2; Length 8220;
Best Local Similarity 54.2%; Pred. No. 0.66;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 2070 ATATGTTATTGGAAATATAGGTAATTTAGTACAGTACATCTTGCATTACATAGGTACT 2129
Db 8067 ATATATATATTTTATATATATATATATATATATATATATATATATATATATATATAT 8126
QY 2130 TCAAGCAACACAAATAAAGTAAATGATAAGTGAACCTTGCTTGTATAGTAATAACA 2189
Db 8127 AAATGAAAAAAGAAAAATGAAATATATAAATAAATAAATAAATAAATAAATAAATAA 8186
QY 2190 AGACCATAGAGAGATAAGTATAGCTAGAGAAAT 2222
Db 8187 AAAAAAAGGAGAAAAATTTTAAAAAATAAT 8219
```

```

RESULT 6
US-08-487-826B-11
; Sequence 11, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-11

Query Match      1.6%; Score 41; DB 2; Length 8220;
Best Local Similarity 54.2%; Pred. No. 0.66;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 2070 ATATGTTATTGGAAATATAGGTAATTTAGTACAGTACATCTTGCATTACATAGGTACT 2129
Db 8067 ATATATATATTTTATATATATATATATATATATATATATATATATATATATATATAT 8126
QY 2130 TCAAGCAACACAAATAAAGTAAATGATAAGTGAACCTTGCTTGTATAGTAATAACA 2189
Db 8127 AAATGAAAAAAGAAAAATGAAATATATAAATAAATAAATAAATAAATAAATAAATAA 8186
QY 2190 AGACCATAGAGAGATAAGTATAGCTAGAGAAAT 2222
Db 8187 AAAAAAAGGAGAAAAATTTTAAAAAATAAT 8219
```

```

RESULT 7
US-09-210-288-11
; Sequence 11, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
```

7  
D  
N  
N  
N  
D  
N  
D  
D  
N  
D

```

;
; NAME/KEY: CDS
; LOCATION: 303..2162
;
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 303..365
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 366..2159
US-08-471-733-5

```

	Query Match	1.6%;	Score 40;	DB 3;	Length 2264;
	Best Local Similarity	46.4%;	Pred. No.	0.61;	
	Matches 130;	Conservative	0;	Mismatches 150;	Indels 0; Gaps 0;
QY	1601	ACTCTTTTCTATPATATTGCTTAGGCTCACCATAGCGAATTCCTCCAGTGTTAAACATTTTC	16600		
Db	328	AATTTATATAAANAATGATTTTTTCATAAATTATACCTTCCTGGTTGTAACTATAAACATTTT	269		
QY	1661	TGTTTTCCACAATTGAACITTTATGGGTTTTGGGGATTTTCTTGTAGTTCCTTATATATATCCCT	1720		
Db	268	ATTAGTTTTTACGTATTTTTTAACGCTCTTAATAATATATATATATAATACATATTACATTATT	209		
QY	1721	ATATATTATATCTATATATTGCCAAAATTTTGACTGTCAGCTACATGTTGGTAAGACACAGGC	1780		
Db	208	GTACATAATAACATGTATACAATATTAAAACTATTAGATAAAAAATTCATATATCATATTGG	149		
QY	1781	AAAGTATTACTGTTAACTAAGTTATTTTTTAAAGTTAAATATATATTTTACGTGCCCTTGGC	1840		
Db	148	TATTTGTGATTTTATACTATTCCAATTTTATAAATAAGATTAAATTGATTGTCAACTAATC	89		
QY	1841	TTTTTTATTCGAGAGTCTACATTTTTTATAGATTCTACATCAG	1880		
Db	88	TTATTTATTTAAATGATTAATTTTATACATAAACCAG	49		

RESULT 10  
US-08-468-878-5/C  
; Sequence 5, Application US/08468878  
; Patent No. 6090586  
; GENERAL INFORMATION:  
; APPLICANT: BERGSTROM SVEN  
; APPLICANT: BARBOUR ALAN G.  
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,878  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/262,220  
; FILING DATE: 20-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: BERGSTROM=3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Borrelia afzelii
; STRAIN: ACAI
; IMMEDIATE SOURCE:
; CLONE: pJB-104
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 303..2162
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 303..365
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 366..2159
US-08-468-878-5

Query Match          1.6%; Score 40; DB 3; Length 2264;
Best Local Similarity 46.4%; Pred. No. 0.61;
Matches 130; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1601 ACTCTTTTCTATATATGCTTAGGCTCACCATAGCGAATTCCTCCAGTGTAAAACTTTC 1660
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 AATTATATAAAATATGATTTTTCATAAATTATACCTTCCTGTTGTAACATAAAACTTTT 269

QY 1661 TGTTTTCACATTTGAACTTTATGGGTTTTGGGGGATTTTCTTGTAGTTCTTATATATCCCT 1720
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 ATTAGTTTTAGTTATTTTTTAAACGCTTAAATAATATATATATATATATATTTACTTATT 209

QY 1721 ATATATTATATCTATATTGC AAAAATTTTGACTGTCAGCTACATGTTGGTAAGACACAGGC 1780
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 208 GTACATAATAACATGTTATACAATATTAAACCTATTAGATAAAAAATTC AATATCATATTGG 149

QY 1781 AAAGTATTACTGTAAGTTATTTTTTAAAGTTAAAAATATATTTTTACGTGCCTTTGGC 1840
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 TATTTGTGATTTTATACTATTCAAATTTTATAAATAAGATTAAATGTATTGTCAACTAATC 89

QY 1841 TTTTTATTGCAGAGTCTACATTTTTTATAGATTCTACATCAG 1880
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 TTATTTATTTTAAATGATTAATTTTATATACTAAATAAACCCAG 49

```

RESULT 11  
US-08-750-494-5/C  
; Sequence 5, Application US/08750494  
; Patent No. 6204018  
; GENERAL INFORMATION:  
; APPLICANT: BERGSTROM SVEN  
; APPLICANT: BARBOUR ALAN G.  
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,494  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/262,220  
; FILING DATE:



ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: BERGSTROM=3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2264 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Borrelia afzelii

STRAIN: ACAI

IMMEDIATE SOURCE:

CLONE: pJB-104

FEATURE:

NAME/KEY: CDS

LOCATION: 303..2162

FEATURE:

NAME/KEY: sig\_peptide

LOCATION: 303..365

FEATURE:

NAME/KEY: mat\_peptide

LOCATION: 366..2159

US-08-750-494-5

Query Match 1.6%; Score 40; DB 4; Length 2264;

Best Local Similarity 46.4%; Pred. No. 0.61;

Matches 130; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1601 ACTCTTTCTATATATGCTTAGGCTCACCATAGCGAATTCCTCCAGTGTAAACCTTTC 1660

Db 328 AATTTATATAAAATATGATTTTTCATAAATTATACCTTCCTGTTGTAACATATAAACTTTT 269

QY 1661 TGTTCACATTTGAACCTTTATGGGTTTGGGGATTTTCTGTAGTTCTTATATATCCCT 1720

Db 268 ATAGTTTAGTTATTTTAAACGCTCTTAATAATATATATATAATACATATTACTTATT 209

QY 1721 ATATATTATCTATATATTGCAAAATTTGACTGTGCTACATGTTGGTAAGACACAGGC 1780

Db 208 GTACATATAACATGTATACATATAATAAACTATTAGATAAAATTCATATCATATTGG 149

QY 1781 AAAGTATTACTGTAACTAAGTTATTTTAAAGTTAAATATATATTTTACGTGCCCTTGGC 1840

Db 148 TATTTGGATTTTATATCTATTTCAAATTTATAAATAAGATTAAATGTATTGTCAACTAATC 89

QY 1841 TTTTATTCAGAGTCTACATTTTATAGATTCTACATCAG 1880

Db 88 TTATTTATTTAAATGATTAATTTATATACTAAATAAACCCAG 49

RESULT 12

US-07-991-867B-32/c

Sequence 32, Application US/07991867B

Patent No. 5476781

GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.

APPLICANT: Hall, Richard L.

APPLICANT: Gruidl, Michael E.

TITLE OF INVENTION: No. 5476781e1 Entomopoxvirus Expression System

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B  
FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO 92/14818

FILING DATE: 12-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/827,685

FILING DATE: 30-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,584

FILING DATE: 19-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF114.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 660 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

US-07-991-867B-32

Query Match 1.5%; Score 39.2; DB 1; Length 660;

Best Local Similarity 48.0%; Pred. No. 0.52;

Matches 142; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

QY 1695 TTTTCTTGTAGTTCTTATATATCCCTATATATATATATATATATATGCTGT 1754

Db 513 TTTTAAATTAATTAATTTTCTATAAAATTTATAATCTTTTATTGTATTGATTCATATT 454

QY 1755 CAGCTACATGTTGGTAAGACACAGGCAAGATTTACTGTAACTAAGTTATTTTAAAGTT 1814

Db 453 TAACTCAACTATATACTAATAAGGAAACATTTATAAGTTACCAAAATTTATTAGATAT 394

QY 1815 AAAATATATTTTACGTGCCTTTGGCTTTTATTGTCAGAGTCTACATTTTATAGATTCTA 1874

Db 393 TATTAATTTTAAATTTTACTAAATTTATAAATAAATTAAGTCA-TTATATTACATG 335

QY 1875 CATCAGATGTTGTCACATTATTTCCATTGGGATGCCATTGTAAGCTGTGATGCGTGT 1934

Db 334 ATTCACAATTTTAAAAATTTCTATAGAATGCGTAGTATAATATTACTATATTGCTATTT 275

QY 1935 TGGAAAAGTGTATCATCATCTAGTTTTTTTTTTCTTCATCTGTTATCATACTTTTAA 1990

Db 274 TGTATAAGATATATCTAAATATGTTATATTTTAAATTTTGTATAAAATTTAAA 219

RESULT 13

US-08-107-755A-32/c

Sequence 32, Application US/08107755A

Patent No. 5721352

GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.

APPLICANT: Hall, Richard L.

APPLICANT: Gruidl, Michael E.

TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF114.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-107-755A-32

Query Match 1.5%; Score 39.2; DB 1; Length 660;  
Best Local Similarity 48.0%; Pred. No. 0.52;  
Matches 142; Conservative 0; Mismatches 153; Indels 1; Gaps 1;  
QY 1695 TTTTCTGTAGTCTTATATATCCCTATATATATATATATATATGCAAAATTTTGACTGT 1754  
Db 513 TTTTAAATTAATTAATTTTCTATAAAATTTATAATCTTTTATTGTTGATTCATAT 454  
QY 1755 CAGCTACATGTTGTAAGACACAGGCAAGATTAATCTGTAAGTAAAGTTATTTTAAAGTT 1814  
Db 453 TAACTCAACTATACTAATAGGAAACATTTATAAGTTACCAAAATTTTATAGATAT 394  
QY 1815 AAAATATATTTTACGTCCTTTGGCTTTTATTTGTCAGAGCTACATTTTATAGATCTA 1874  
Db 393 TATTAATTTTAAATTTACTAAATTAATAAAATTAAGTCA-TTTATATTACATG 335  
QY 1875 CATCAGATGTTGTCACCTTATTTCCATGGGATTCCTATGTAAGCTGTGTATGTCGCTGT 1934  
Db 334 ATTCACAATTTAAATAATCTATAGAAATGTTAGTATAATAATTTATTTGCTATTTT 275  
QY 1935 TGGAAAAGTATTCATCTAGTTTCTTTCTTCTCATCTGTTATCATCTTTTAA 1990  
Db 274 TGTATAAGATATATCTAAATATGTTATATTTTAAATTTTCTTATAAAATTTTAA 219

RESULT 14  
US-08-544-332-32/c  
Sequence 32, Application US/08544332  
Patent No. 5935777  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gerard H. Bencen  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville

STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/544,332  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,867  
FILING DATE: 07-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/107,755  
FILING DATE: 19-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bencen, Gerard H.  
REGISTRATION NUMBER: 35,746  
REFERENCE/DOCKET NUMBER: UF114.C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-544-332-32

Query Match 1.5%; Score 39.2; DB 2; Length 660;  
Best Local Similarity 48.0%; Pred. No. 0.52;  
Matches 142; Conservative 0; Mismatches 153; Indels 1; Gaps 1;  
QY 1695 TTTTCTGTAGTCTTATATATCCCTATATATATATATATATGCAAAATTTTGACTGT 1754  
Db 513 TTTTAAATTAATTAATTTTCTATAAAATTTATAATCTTTTATTGTTGATTCATAT 454  
QY 1755 CAGCTACATGTTGTAAGACACAGGCAAGATTAATCTGTAAGTAAAGTTATTTTAAAGTT 1814  
Db 453 TAACTCAACTATACTAATAGGAAACATTTATAAGTTACCAAAATTTTATAGATAT 394  
QY 1815 AAAATATATTTTACGTCCTTTGGCTTTTATTTGTCAGAGCTACATTTTATAGATCTA 1874  
Db 393 TATTAATTTTAAATTTACTAAATTAATAAAATTAAGTCA-TTTATATTACATG 335  
QY 1875 CATCAGATGTTGTCACCTTATTTCCATGGGATTCCTATGTAAGCTGTGTATGTCGCTGT 1934  
Db 334 ATTCACAATTTAAATAATCTATAGAAATGTTAGTATAATAATTTATTTGCTATTTT 275  
QY 1935 TGGAAAAGTATTCATCTAGTTTCTTTCTTCTCATCTGTTATCATCTTTTAA 1990  
Db 274 TGTATAAGATATATCTAAATATGTTATATTTTAAATTTTCTTATAAAATTTTAA 219

RESULT 15  
US-09-370-861A-32/c  
Sequence 32, Application US/09370861A  
Patent No. 6410221  
GENERAL INFORMATION:

Search completed: July 25, 2003, 19:56:50  
Job time : 127.381 secs





us-09-806-276a-4.rng

Tue Jul 29 09:34:15 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 13:53:40 ; Search time 587.148 Seconds  
(without alignments)  
9765.155 Million cell updates/sec

Title: US-09-806-276A-4  
Perfect score: 2546  
Sequence: 1 cggggctagccggagacc.....cagctanaacgagggcagta 2546

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2545	100.0	2546	21 AAA09155	Human BMDSP-2 codi
2	491.8	19.3	564	24 ABK64131	Human benign prost
3	330.8	13.0	1617	23 AAS89293	DNA encoding novel
4	133.8	5.3	156	21 AAZ94524	Cartilage-associat
5	124.8	4.9	1079	24 ABQ17530	Oligonucleotide fo
6	124.8	4.9	1079	24 ABQ17531	Oligonucleotide fo
7	99.2	3.9	1079	24 ABQ17532	Oligonucleotide fo
8	99.2	3.9	1079	24 ABQ17533	Oligonucleotide fo
9	85	3.3	413	22 ABA46549	Human breast cell

10	85	3.3	413	22	ABA64411	Human foetal liver
11	85	3.3	413	22	AAK12870	Human brain expres
12	85	3.3	413	22	AAI19399	Probe #9332 for ge
13	85	3.3	413	22	AAI44587	Probe #13273 used
14	85	3.3	413	22	AAI05124	Probe #5115 used t
15	85	3.3	413	24	ABSI2665	Human genome-deriv
16	85	3.3	413	24	ABA51732	Human foetal liver
17	85	3.3	467	22	AAK00042	Human brain expres
18	85	3.3	467	22	AAI10103	Probe #36 for gene
19	85	3.3	467	22	AAI31349	Probe #35 used to
20	85	3.3	467	22	AAI00044	Probe #35 used to
21	85	3.3	467	24	ABS00045	Human genome-deriv
22	64.4	2.5	185	21	AAZ94497	Cartilage-associat
23	62	2.4	11812	22	AAS45502	Chemically pretrea
24	62	2.4	11812	22	AAS46742	Tumour suppressor
25	62	2.4	11812	24	ABL34119	Human immune syste
26	62	2.4	11812	24	ABK28432	DNA transcription
27	60	2.4	60	24	ABN41082	Human spliced tran
28	59	2.3	8079	24	ABL92313	Chemically treated
29	58.2	2.3	6963	24	ABL32978	Human immune syste
30	57.8	2.3	7571	24	ABL32527	Human immune syste
31	57.2	2.2	6222	24	ABL32693	Human immune syste
32	57.2	2.2	6285	24	ABL33497	Human anglogenesis
33	56.8	2.2	6070	24	ABQ67130	Chemically treated
34	56.8	2.2	6070	24	ABL70372	Human immune syste
35	56.8	2.2	6070	24	ABL33679	Human metastasis a
36	56.8	2.2	6070	24	ABL34579	Chemically pretrea
37	56.8	2.2	9539	24	ABK28180	DNA transcription
38	56.8	2.2	9539	24	ABK28180	Human immune syste
39	56.2	2.2	14006	24	ABL33958	Human chemically p
40	55.6	2.2	7441	24	ABK40058	Chemically pretrea
41	55.6	2.2	7479	24	AAS63345	AmEPV genome fragm
42	55.6	2.2	50000	24	ABL55643	Human immune syste
43	55.4	2.2	5981	24	ABL33763	Human immune syste
44	55.2	2.2	14950	24	ABL33257	Human immune syste
45	55.2	2.2	15373	24	ABL32466	Human immune syste

ALIGNMENTS

RESULT 1  
AAA09155  
ID AAA09155 standard; DNA; 2546 BP.

XX  
AC AAA09155;

XX  
DT 10-AUG-2000 (first entry)

XX  
DE Human BMDSP-2 coding sequence.

XX  
KW Bone marrow-derived serum protein; BMDSP-2; cytostatic; anti-HIV;  
antiasthmatic; antiatherosclerotic; anti-inflammatory; antiarthritic;  
antisclerotic; hypotensive; antiviral; antiparasitic; ss.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT CDS 413..1177

FT FT /\*tag= a

FT FT /product= BMDSP-2

XX  
PN WO200020588-A2.

XX  
PD 13-APR-2000.

XX  
PF 01-OCT-1999; 99WO-US22908.

XX  
PR 02-OCT-1998; 98US-0165621.

XX  
PA (INCY-) INCYTE PHARM INC.

XX  
PI Tang YT, Corley NC, Guegler KJ, Lu DAM;

XX WPI; 2000-303775/26.  
DR P-PSDB; AAY92240.  
XX  
PT Purified polypeptide for treating or preventing disorders associated  
PT with decreased expression or activity of bone marrow-derived serum  
PT proteins  
XX  
PS Claim 9; Page 70-71; 72pp; English.  
XX  
CC Human bone marrow-derived serum proteins (BMDSP) 1 and BMDSP-2 are  
CC useful for treating or preventing a disorder associated with decreased  
CC expression or activity of bone marrow-derived serum proteins.  
CC Antagonists of BMDSP are useful for treating or preventing a disorder  
CC associated with increased expression or activity of bone marrow-derived  
CC serum proteins. The disorders include cancers (melanoma, adenocarcinoma,  
CC sarcoma), immune disorders (acquired immunodeficiency syndrome (AIDS),  
CC asthma, atherosclerosis, Crohn's disease, bronchitis, multiple sclerosis,  
CC osteo- and rheumatoid arthritis), viral infections, parasitic infections  
CC (schistosoma, tapeworm), and vascular disorders (arteriosclerosis,  
XX hypertension, vasculitis).  
SQ Sequence 2546 BP; 708 A; 576 C; 553 G; 708 T; 1 other;  
  
Query Match 100.0%; Score 2545; DB 21; Length 2546;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGCGGCTGGGGGCGCTTCACGCCGTCTCGGAGCGG 60  
DB |||||  
QY 61 ATAATGCGGTGAGCAGGACCCACCGCGGAGACTCGGCTGGATCTGCGACAGCGGCGAGG 120  
DB |||||  
QY 121 GATTGCGTGGCGCGGGAGGCGCGGCGGCGGCTGGGATCCTCAGCGCGGCGCGGT 180  
DB |||||  
QY 181 TGTCTGTTGTGGTCAAGACTGGATGATGTAAGTGGCTCTAGGAAGCCTCACTTGGC 240  
DB |||||  
QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGGAAGCCACTTCTGAAGCCAC 300  
DB |||||  
QY 301 TTGAGAAAATGATGTGACAGTTCTCTATCAAAAGGATTCAGAAACATATACCATCTGTG 360  
DB |||||  
QY 361 AAGAAAGTGGCCCTTCTCCCGCTTGCAAAATAGACATTTCAAAATCCAAATGCCAGC 420  
DB |||||  
QY 421 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAAAGAAAGGAAAGAAATTTAAACTGAG 480  
DB |||||  
QY 481 GGACATTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
DB |||||  
QY 541 CATTTGGCAAAGGGGCCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB |||||  
QY 601 CGAGCTTTTACCTGGAAACAGGAGAAAGCACACCTGGGCGAGTTCCCTGGGCATATGA 660  
DB |||||  
QY 661 GTTCTTCCGGCCAAACAGCACCTCGGACTCTGTGTACAGAAACGCCCTCCCGGCTGCT 720  
DB |||||

Db 661 GTTCTTCCGGCCAAACAGCACCTCGGACTCTGTGTTACAGAAACGCCCTCCCGGCTGCT 720  
QY 721 CAAAAATGCCATCTCCCTCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATT 780  
Db |||||  
QY 781 GTCACCAAGTGCATTTAAATTTCCAAACAGGAGTCTTTCGGGCCAGCAAGCTGCCAGGCT 840  
Db |||||  
QY 841 TAGCTGCGAGCCGCTCATGGAGGAAAAAGCTCAGGAGAAAAAGCTGTTGGAGAAATGG 900  
Db |||||  
QY 901 GACAGTCCACAGGGGAGACACCTCTGTTGGGCTCCAGCGGTTCTGCATCTCAGTCCAGCCA 960  
Db |||||  
QY 961 AGGCAGAGACAGCCACTCTCCAGCCTGTCCGAACAGTACCCCGACTGGCCAGCGGAGGA 1020  
Db |||||  
QY 1021 CATGTTTGACCATCCCAACCCCATCGGAGCTCATCAAGGAAAGACTAAGTCAGAGGAGTC 1080  
Db |||||  
QY 1081 CCTCTCTGACCTTACAGGTTCCCTCTCTCCCTGCGAGCTTGAATCTTGGGCCCTCACTTTT 1140  
Db |||||  
QY 1141 GGATGAGGTGCTGAATGTAATGGATAAAATAAGTAACAAGATGCCAACTTTTTCCTTT 1200  
Db |||||  
QY 1201 GGGTAAAGGTACAAAAACAACTAACACAGTTGAAGAGAAAGGCTTCCGGAGCTGA 1260  
Db |||||  
QY 1261 TTTGACAGTTTGTGTTGGGTTTCTAAATAATATTTCTTACAAAGTATTTTACCTGT 1320  
Db |||||  
QY 1321 TATGCCCCGTTTGCAAAAACAATTTAGAAAAAAACAACAAGCAAAACCTATCTGGCAA 1380  
Db |||||  
QY 1381 AAAAGGAAGTGAAGTCAAGAGCCCATTTTTCAGGAGGCTTGGTGTGTTCCGCTCACATAT 1440  
Db |||||  
QY 1441 TGTGTCAGACACACAAGAAATCTGGCTTGGCCAGGATGGCAGTATGAAGGGCTG 1500  
Db |||||  
QY 1501 AGCGAGTCAATTAAGGAATTCACGGAATTTTATAGCACTCCGACATTTCTGAGCAAG 1560  
Db |||||  
QY 1561 AGGAGTCAAAATTTATTAACACCTAAGCCCTTTTGTAGACTCTTTCTATATATGCT 1620  
Db |||||  
QY 1621 TAGGCTCACCATAGCGAATTTCTCCAGTGTAAAAACTTTTCTGTTTTCACATTTGAACCTT 1680  
Db |||||  
QY 1681 ATGGGTTTGGGATTTCTGTAGTTCTTATATATATCCCTATATATATATATATGCT 1740  
Db |||||  
QY 1741 AAAATTTGACTGACGTACATGTTGGTGAAGACACAGGCAAGTATTAAGTAACTAAG 1800  
Db |||||

Db 1741 AAAATTTTGACTGCTACATGTTGGTAAGACACAGGCAAACTATTACTGTAACTAAG 1800  
QY 1801 TTATTTTAAAGTTAAATATATTTTACGTCCTTTGGCTTTTATTCAGAGCTCTACA 1860  
Db 1801 TTTATTTTAAAGTTAAATATATTTTACGTCCTTTGGCTTTTATTCAGAGCTCTACA 1860  
QY 1861 TTTTATAGATTCTACATCAGATGTTGTCACTTATTTCCATTGGGATTCATTTGTAAGCTG 1920  
Db 1861 TTTTATAGATTCTACATCAGATGTTGTCACTTATTTCCATTGGGATTCATTTGTAAGCTG 1920  
QY 1921 TGTATGTCGCTGTTGGAAAAGTGTATTCATACCTTATTTTCTTCTCATCTGTTATC 1980  
Db 1921 TGTATGTCGCTGTTGGAAAAGTGTATTCATACCTTATTTTCTTCTCATCTGTTATC 1980  
QY 1981 ATACTTTTAAACAGCAACCAATAAACGGATTGTAAAGTGTAAGGACAGAGTTACTCATGAT 2040  
Db 1981 ATACTTTTAAACAGCAACCAATAAACGGATTGTAAAGTGTAAGGACAGAGTTACTCATGAT 2040  
QY 2041 GCTTCTGCAGAGACTGGGGCTACACCATATGTTATTTGGAAATATAGGTATTTTAGT 2100  
Db 2041 GCTTCTGCAGAGACTGGGGCTACACCATATGTTATTTGGAAATATAGGTATTTTAGT 2100  
QY 2101 ACAGTACATACCTTGCATTACATAGGTACTTCAAGCAACACAAATAAAAGTAAATGATAAA 2160  
Db 2101 ACAGTACATACCTTGCATTACATAGGTACTTCAAGCAACACAAATAAAAGTAAATGATAAA 2160  
QY 2161 GTGAACCTTGCTTGTATAGTAATAAACAGACCATAGAGATAAGTATAGCTAGAGAA 2220  
Db 2161 GTGAACCTTGCTTGTATAGTAATAAACAGACCATAGAGATAAGTATAGCTAGAGAA 2220  
QY 2221 ATTGCTTCTCTGAAATGTACATGAGCCCTTAAGTAAGAGATGATTTCCATCTACTCTCA 2280  
Db 2221 ATTGCTTCTCTGAAATGTACATGAGCCCTTAAGTAAGAGATGATTTCCATCTACTCTCA 2280  
QY 2281 TTTTGATTAATCTCCTTAATGCTTGTAGAGGCTAGAACTGAGCCTCTCTACTTTTGGAAAA 2340  
Db 2281 TTTTGATTAATCTCCTTAATGCTTGTAGAGGCTAGAACTGAGCCTCTCTACTTTTGGAAAA 2340  
QY 2341 ATGAACATGTGAGGTGAGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2400  
Db 2341 ATGAACATGTGAGGTGAGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2400  
QY 2401 CAGTGTCTATCTCTGAGCATCTCTGACTTGAACACCTTCTACAGCAAACTCTTGTCAAG 2460  
Db 2401 CAGTGTCTATCTCTGAGCATCTCTGACTTGAACACCTTCTACAGCAAACTCTTGTCAAG 2460  
QY 2461 TCCAGTTTCATCCCTGTAGGCAAAATGCTTTTTCACGAGAAAGTGCCATATAGACGAGA 2520  
Db 2461 TCCAGTTTCATCCCTGTAGGCAAAATGCTTTTTCACGAGAAAGTGCCATATAGACGAGA 2520  
QY 2521 TAAAGCAGCTANAAACGAGGCGAGTA 2546  
Db 2521 TAAAGCAGCTANAAACGAGGCGAGTA 2546

RESULT 2  
ABK64131/c  
ID ABK64131 standard; DNA; 564 BP.

XX  
AC ABK64131;  
XX  
DT 18-JUN-2002 (first entry)

DE Human benign prostatic hyperplasia gene #26.

XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.

OS Homo sapiens.

XX WO200212440-A2.

PN 14-FEB-2002.

XX

PF 07-AUG-2001; 2001WO-US24708.  
XX  
PR 07-AUG-2000; 2000US-223323P.  
PR 05-JUN-2001; 2001US-0873319.  
XX  
PA (GENE-) GENE LOGIC INC.  
PA (NISB ) JAPAN TOBACCO INC.  
XX  
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
XX  
XX WPI; 2002-257476/30.  
DR  
XX  
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
PT detecting expression levels of one or more genes in prostate cells from  
PT patient that are differentially regulated compared to normal prostate  
PT cells  
XX  
PS Disclosure; Page 83; 444pp; English.  
XX  
CC The invention relates to a method of diagnosing (I) the onset or  
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
CC or identifying an agent that modulates the onset or progression of BPH.  
CC The method is based on changes in gene expression in BPH tissue isolated  
CC from patients exhibiting different clinical states of prostate  
CC hyperplasia as compared to normal prostate tissue. (I) comprises  
CC detecting the expression levels of one or more genes in prostate cells  
CC from the subject that are differentially regulated compared to normal  
CC prostate cells. (II) comprises preparing a first gene expression profile  
CC of BPH cells or BPH-like cell population, exposing the cells to the  
CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles.  
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is  
CC useful for identifying an agent that modulates the onset or progression  
CC of BPH. The methods are useful to present information identifying  
CC the expression level in a tissue or cells, by comparing the expression  
CC level of genes given in the specification in the tissue or cells to the  
CC level of expression of gene in the database, and displaying the  
CC expression levels of at least one gene in the tissue or cell sample  
CC compared to the expression level in BPH. Agents using (II) are useful for  
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human  
CC benign prostatic hyperplasia gene sequences of the invention.  
XX  
SQ Sequence 564 BP; 210 A; 93 C; 85 G; 174 T; 2 other;

Query Match 19.3%; Score 491.8; DB 24; Length 564;  
Best Local Similarity 98.6%; Pred. No. 2.1e-118;  
Matches 506; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1653 AACCTTTCTGTTTTCACATTTGAACCTTATGGGTTTGGGATTTCTTGTAGTTCTTAT 1712  
Db 523 AAACCTTTCTGTTTTCACATTTGAACCTTATGGGTTTGGGATTTCTTGTAGTTCTTAT 464  
QY 1713 ATATCCCTATATATATATCTATATATGCAAAATTTTGAAGTAAAGTAAAGTAAAGTAAAG 1772  
Db 463 ATATCCCTATATATATATCTATATATGCAAAATTTTGAAGTAAAGTAAAGTAAAGTAAAG 404  
QY 1773 ACACAGGCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1832  
Db 403 ACACAGGCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 345  
QY 1833 CCTTTGGCTTTTATTCGAGAGTCTACATTTATAGATTTCTACATCAGATGTTGTCACCT 1892  
Db 344 CCTTTGGCTTTTATTCGAGAGTCTACATTTATAGATTTCTACATCAGATGTTGTCACCT 285  
QY 1893 ATTTCCATTGGGATTCCTATTTGAAGCTGTATGTCGCTGTTTGGAAAGTGTATTTCATA 1952  
Db 284 ATTTCCATTGGGATTCCTATTTGAAGCTGTATGTCGCTGTTTGGAAAGTGTATTTCATA 225  
QY 1953 CTAGTATTTTCTTCTCATCTGTTATCATACCTTTTAAACAGCAACCAATTAACGGATTGTA 2012  
Db 224 CTAGTATTTTCTTCTCATCTGTTATCATACCTTTTAAACAGCAACCAATTAACGGATTGTA 165.  
QY 2013 AAGTGTAAGGCAACAGCTTACTCATGATGCTTCTGCAGAGACTGTGGGCTACACACATA 2072







us-09-806-276a-4.rng

Tue Jul 29 09:34:15 2003

articular chondrocyte differentiation. They correspond to portions of genes that are differentially regulated in human articular chondrocytes 48 hr after transfer to an alginate suspension culture as compared to those expressed in dedifferentiated monolayer culture. The polynucleotides can be used as markers to identify chondrocytes, as probes for detecting chondrocytes, or can be incorporated into vectors for recombinant production of chondrocyte differentiation-related polypeptides. Such polypeptides are useful for screening for modulator compounds and for production of antibodies used for in vivo imaging or immunoassay.

Sequence 156 BP; 37 A; 26 C; 27 G; 57 T; 9 other;  
Query Match 5.3%; Score 133.8; DB 21; Length 156;  
Best Local Similarity 91.1%; Pred. No. 5.7e-25;  
Matches 144; Conservative 6; Mismatches 6; Indels 2; Gaps 1;  
QY 2239 ACATGAGCCCTTAAGGTAAGAGATGATTTCCATCTACTCTCATTTGATTACTTCTTAT 2298  
Db 1 ACATGAGCCCTTAAGGTAAGAGATGATTTNNATCTACTCTCATTTGATTACTTCTTAT 60  
QY 2299 GGTGAGAGGCTAGAACTGAGCCTCTCTACTTTTGGAAAAATGACATGTGAGGTGAG 2358  
Db 61 GGTGAGAGGCTAGAACTGAGCCTCTCTACTTTGK--RAATGACATGTGAGGTGAG 118  
QY 2359 ATTTTCTTTTCTTTTAAAGTCACTGATGATGATGATGATGATGATGATGATGATGAT 2396  
Db 119 ATTTTCTTTTCTTTTAAAGTCACTGATGATGATGATGATGATGATGATGATGATGAT 156

RESULT 5  
ABQ17530 standard; DNA; 1079 BP.  
XX  
AC ABQ17530;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4121.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
DR  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic

DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 1079 BP; 140 A; 124 C; 384 G; 431 T; 0 other;  
Query Match 4.9%; Score 124.8; DB 24; Length 1079;  
Best Local Similarity 78.1%; Pred. No. 3.7e-22;  
Matches 150; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 1 CGGGGCTAGCCCGGAGACCCGGCCACCGGCGCTGGGGCGCTTACGCCGCTCTCGGAGCGG 60  
Db 724 CGGGGCTAGCTCGGAGATTCGGTTATCGGTTTACGTTTGGGGCGGTTTTCGGAGCGG 783  
QY 61 ATAATCGGTGAGCAGGACACGACCGCGGAGGCTGGGCTGGATCTGCGCAGCGGCGAGG 120  
Db 784 ATAATCGGTGAGTAGTATTACGTGCGGTAGATTGCGTTGGATTTGCGTATAGCGGTAGG 843  
QY 121 GATTGCGTGGCCCGCGGAGGCGCGGCGGAGCGGCTGGGATCTTACGCCGCGGCGGTT 180  
Db 844 GATTGCGTGGCTCGCGGAGGTTTCGGGGTAGCGGTTGGGATTTTACGCGGCGGTTG 903  
QY 181 TGTCTCTGGTTGT 192  
Db 904 AGTCGGGGTAGT 915

RESULT 6  
ABQ17531/c  
ID ABQ17531 standard; DNA; 1079 BP.  
XX  
AC ABQ17531;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4122.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
DR  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT



Tue Jul 29 09:34:15 2003

01-FEB-2002 (first entry)  
Human breast cell single exon nucleic acid probe #5244.  
Human; microarray; single exon probe; gene expression; breast;  
disease; cancer; ss.  
Homo sapiens.  
WO200157271-A2.  
09-AUG-2001.  
30-JAN-2001; 2001WO-US00662.  
04-FEB-2000; 2000US-0180312.  
26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408.  
03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-0234687.  
27-SEP-2000; 2000US-0236359.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-496933/54.  
New spatially-addressable set of single exon nucleic acid probes,  
useful for measuring gene expression in sample derived from human  
breast, comprises number of single exon nucleic acid probes -  
Claim 4; SEQ ID NO 5244; 327pp + sequence listing; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human breast and BT 474 cells. The method involves contacting  
the probes with a collection of detectably labelled nucleic acids  
derived from mRNA of human breast, and then measuring the label  
bound to each probe of the microarray. The probes are useful for  
verifying the expression of regions of genomic DNA predicted to  
encode proteins. They are useful for gene discovery, and for  
determining predisposition and/or prognosing the toxicity of chemical  
agents on cells. The microarray of this invention presents a far greater  
diversity of probes for measuring gene expression, with far less bias  
than expressed sequence tag microarrays. The method is suitable for  
rapid production of functional information from genomic sequence. The  
present sequence is a single exon nucleic acid probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 413 BP; 73 A; 131 C; 131 G; 78 T; 0 other;  
Query Match 3.3%; Score 85; DB 22; Length 413;  
Best Local Similarity 64.7%; Pred. No. 5.8e-12;  
Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;  
QY 409 CAAAATGCCAGCCAGACCCCAATTTACTGAAAGCAGCCAATAACAAAGAAAGAA 468  
Db 101 CACCATGTCCACCAAGTGTGCTCTATCTGAAAGCGTGGCAGTCGCAAG--GGCAAGAA 157  
QY 469 ATTTAAACTGAGGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 528  
Db 158 GGAGAGGCTTCGGGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 217  
QY 529 CCACACCATCCACATTGGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTCTCTTCT 588  
Db 218 CCACACCATTCATATTGGCAGTGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 277  
QY 589 TCAAGGGAACACTACGAGCTTTTACCTGGAAACCCAGGAGAAAG 629

01-FEB-2002 (first entry)  
Human breast cell single exon nucleic acid probe #5244.  
Human; microarray; single exon probe; gene expression; breast;  
disease; cancer; ss.  
Homo sapiens.  
WO200157271-A2.  
09-AUG-2001.  
30-JAN-2001; 2001WO-US00662.  
04-FEB-2000; 2000US-0180312.  
26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408.  
03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-0234687.  
27-SEP-2000; 2000US-0236359.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-496933/54.  
New spatially-addressable set of single exon nucleic acid probes,  
useful for measuring gene expression in sample derived from human  
breast, comprises number of single exon nucleic acid probes -  
Claim 4; SEQ ID NO 5244; 327pp + sequence listing; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human breast and BT 474 cells. The method involves contacting  
the probes with a collection of detectably labelled nucleic acids  
derived from mRNA of human breast, and then measuring the label  
bound to each probe of the microarray. The probes are useful for  
verifying the expression of regions of genomic DNA predicted to  
encode proteins. They are useful for gene discovery, and for  
determining predisposition and/or prognosing the toxicity of chemical  
agents on cells. The microarray of this invention presents a far greater  
diversity of probes for measuring gene expression, with far less bias  
than expressed sequence tag microarrays. The method is suitable for  
rapid production of functional information from genomic sequence. The  
present sequence is a single exon nucleic acid probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 413 BP; 73 A; 131 C; 131 G; 78 T; 0 other;  
Query Match 3.3%; Score 85; DB 22; Length 413;  
Best Local Similarity 64.7%; Pred. No. 5.8e-12;  
Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;  
QY 409 CAAAATGCCAGCCAGACCCCAATTTACTGAAAGCAGCCAATAACAAAGAAAGAA 468  
Db 101 CACCATGTCCACCAAGTGTGCTCTATCTGAAAGCGTGGCAGTCGCAAG--GGCAAGAA 157  
QY 469 ATTTAAACTGAGGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 528  
Db 158 GGAGAGGCTTCGGGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 217  
QY 529 CCACACCATCCACATTGGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTCTCTTCT 588  
Db 218 CCACACCATTCATATTGGCAGTGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 277  
QY 589 TCAAGGGAACACTACGAGCTTTTACCTGGAAACCCAGGAGAAAG 629



Db 278 GCAGGCAAGTTCCACCTCCTGCGGGGACCACATGGTGGAGG 318

## RESULT 10

ABA64411

ID ABA64411 standard; DNA; 413 BP.

XX AC ABA64411;

XX DT 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #12716.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human fetal liver

XX PS Claim 4; SEQ ID NO 12716; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for

XX CC measuring human gene expression in a sample derived from human foetal

XX CC liver. The single exon nucleic acid probes may be used for predicting,

XX CC measuring and displaying gene expression in samples derived from human

XX CC foetal liver. The present sequence is a single exon nucleic acid

XX CC probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 413 BP; 73 A; 131 C; 131 G; 78 T; 0 other;

Query Match

Best Local Similarity 3.3%; Score 85; DB 22; Length 413;

Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 409 CAAATGCCAGCCAGACCCCAATTTACCTGAAGCAGCCCAATGAAGAAAGAAAGAA 468

Db 101 CACCATGTCCACCAAGGTGCCATCTATCTGAAGCGTGGCAGTGGCAAG---GGCAAGAA 157

QY 469 ATTTAACTGAGGACATTTCTCTCTCTGATATGATCATGTCCTCCGCTTGGAGACTTTCG 528

Db 158 GGAGAAGCTTCGGACCTGCTGCTCTCGGACATGATCAGCCACCGCTGGGGACTTCCG 217

QY 529 CCACACCATCCACATTTGGCAAGAGGGCCAGCAGATGTCCTTTGGAGATATTTCTTCT 588

Db 218 CCACACCATTCATATTTGGCAGTGGCGGGCGGAGTGCATGTTTGGCGACATCTCCTTCT 277

QY 589 TCAAGGGAACACTACGAGCTTTTACCTGGAAACAGGAGAAAG 629

Db 278 GCAGGCAAGTTCCACCTCCTGCGGGGACCACATGGTGGAGG 318

Db 278 GCAGGCAAGTTCCACCTCCTGCGGGGACCACATGGTGGAGG 318

## RESULT 11

AAK12870

ID AAK12870 standard; DNA; 413 BP.

XX AC AAK12870;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 12861.

XX KW Human; brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human

XX PT brains

XX PS Example 4; SEQ ID NO: 12861; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid

XX CC probes which are derived from genomic sequences expressed in the human

XX CC brain. They can be used to measure gene expression in brain cell samples,

XX CC which may enable the diagnosis and improved treatment of nervous system

XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX CC epilepsy and cancers. The present sequence is one of the probes of the

XX CC invention.

XX SQ Sequence 413 BP; 73 A; 131 C; 131 G; 78 T; 0 other;

Query Match

Best Local Similarity 3.3%; Score 85; DB 22; Length 413;

Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 409 CAAATGCCAGCCAGACCCCAATTTACCTGAAGCAGCCCAATGAAGAAAGAAAGAA 468

Db 101 CACCATGTCCACCAAGGTGCCATCTATCTGAAGCGTGGCAGTGGCAAG---GGCAAGAA 157

QY 469 ATTTAACTGAGGACATTTCTCTCTCTGATATGATCATGTCCTCCGCTTGGAGACTTTCG 528

Db 158 GGAGAAGCTTCGGACCTGCTGCTCTCGGACATGATCAGCCACCGCTGGGGACTTCCG 217

QY 529 CCACACCATCCACATTTGGCAAGAGGGCCAGCAGATGTCCTTTGGAGATATTTCTTCT 588

Db 218 CCACACCATTCATATTTGGCAGTGGCGGGCGGAGTGCATGTTTGGCGACATCTCCTTCT 277

QY 589 TCAAGGGAACACTACGAGCTTTTACCTGGAAACAGGAGAAAG 629

Db 278 GCAGGCAAGTTCCACCTCCTGCGGGGACCACATGGTGGAGG 318





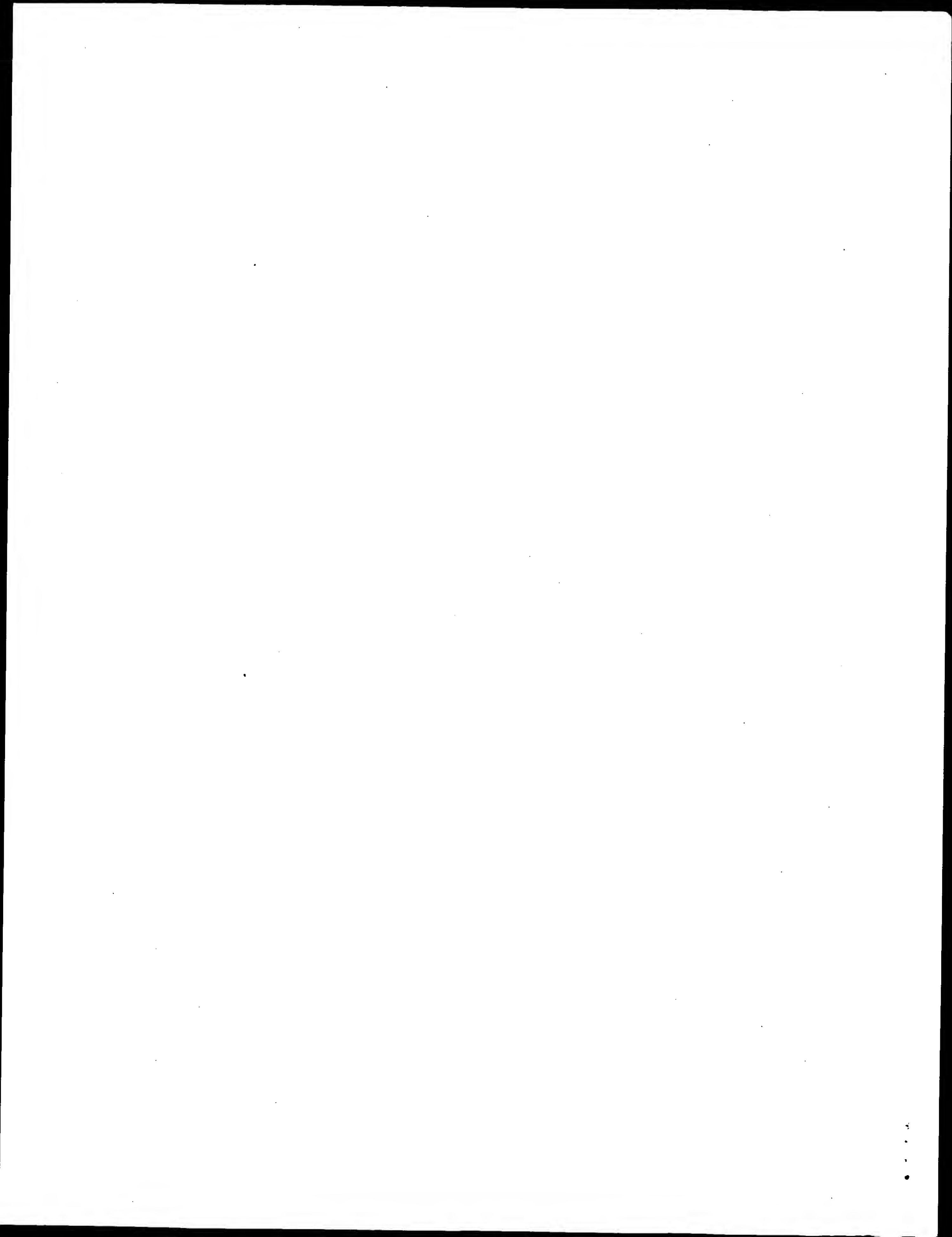
```
AAI05124
ID AAI05124 standard; DNA; 413 BP.
XX
AC AAI05124;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #5115 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 5115; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 413 BP; 73 A; 131 C; 131 G; 78 T; 0 other;

Query Match 3.3%; Score 85; DB 22; Length 413;
Best Local Similarity 64.7%; Pred. No. 5.8e-12;
Matches 143; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 409 CAAATGCGCAGCCCAAGACCCCAATTTACCTGAAAGCGGCAATACAAAGAAAGAAAGAA 468
DB || || || || || || || || || || || || || || || || || || || || || ||
101 CACCATGCTCCACCAAGGTGCCATCTATCTGAAAGCGTGGCAGTCGCAAG---GGCAAGAA 157
QY 469 ATTTAACTGAGGGACATTTCTGTCCTCTGATATGATGATGATGATGATGATGATGATGAT 528
DB || || || || || || || || || || || || || || || || || || || || || ||
158 GGAGAGCTTCGGGACCTGCTGCTCTCGGACATGATGATGATGATGATGATGATGATGATGAT 217
QY 529 CCACACCATCCACATTTGGCAAGAGGGCCAGCACGATGCTTTTGGAGATATTTCTTCT 588
DB || || || || || || || || || || || || || || || || || || || || || ||
218 CCACACCATTCATATTGGCAGTGGGGCGGCGAGTGACATGTTTGGCGACATCTCCTCCT 277
QY 589 TCAAGGGAAGTACGAGCTTTTACCTGGAAACAGGAGAAAG 629
DB || || || || || || || || || || || || || || || || || || || || || ||
```

```
Db 278 GCAGGGCAAGTTCCACCTCCTCCGGGGACCATGGTGGAGG 318
RESULT 15
ABS12665
ID ABS12665 standard; DNA; 413 BP.
XX
AC ABS12665;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 12656.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 4; SEQ ID No 12656; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
```

Search completed: July 25, 2003, 15:52:53  
Job time : 591.148 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 14:39:25 ; Search time 1425.45 Seconds  
(without alignments)  
10929.928 Million cell updates/sec

Title: US-09-806-276A-3  
Perfect score: 962  
Sequence: 1 tcgagccgattcgctcgag.....gaatctttgcaaaaaaaaaa 962

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	803.2	83.5	962	9	AL554570
2	787.6	81.9	897	14	BQ707991
3	787.6	81.9	923	14	BQ709678
4	783.2	81.4	960	14	BQ711479
5	782.8	81.4	888	14	BQ710002
6	771	80.1	972	14	BQ706417

C	7	768	79.8	926	14	BQ712054	BQ712054	AGENCOURT
	8	765.8	79.6	924	12	BG745886	BG745886	602724056
	9	764.2	79.4	931	14	BQ707407	BQ707407	AGENCOURT
	10	758.2	78.8	911	14	BQ899173	BQ899173	AGENCOURT
	11	754.4	78.4	974	14	BQ707037	BQ707037	AGENCOURT
	12	747.2	77.7	950	14	BQ709764	BQ709764	AGENCOURT
	13	743	77.2	888	14	BQ706903	BQ706903	AGENCOURT
	14	738.4	76.8	913	14	BQ711070	BQ711070	AGENCOURT
	15	732	76.1	869	12	BG754043	BG754043	602709571
	16	728.4	75.7	938	14	BQ709175	BQ709175	AGENCOURT
	17	725.6	75.4	841	12	BG758779	BG758779	602713139
	18	723.2	75.2	914	14	BQ707157	BQ707157	AGENCOURT
	19	720.6	74.9	923	14	BQ711051	BQ711051	AGENCOURT
	20	717.8	74.6	928	14	BQ709024	BQ709024	AGENCOURT
	21	717.6	74.6	881	14	BQ709375	BQ709375	AGENCOURT
	22	714	74.2	956	14	BQ706941	BQ706941	AGENCOURT
	23	713.4	74.2	1007	14	BM914510	BM914510	AGENCOURT
	24	708.4	73.6	840	12	BG685798	BG685798	602637907
	25	708	73.6	849	13	BM007345	BM007345	603616565
	26	707	73.5	917	14	BQ711815	BQ711815	AGENCOURT
	27	706.8	73.5	852	12	BG756918	BG756918	602715417
	28	706.4	73.4	975	14	BQ707855	BQ707855	AGENCOURT
	29	704	73.2	851	12	BG754860	BG754860	602714247
	30	703.6	73.1	988	12	BG756209	BG756209	602713519
	31	703.4	73.1	893	12	BG397512	BG397512	602439341
	32	699.6	72.7	1075	12	BG758924	BG758924	602713036
	33	699	72.7	1010	14	BQ711412	BQ711412	AGENCOURT
	34	697.8	72.5	924	12	BG745784	BG745784	602723920
	35	697	72.5	938	14	BQ709467	BQ709467	AGENCOURT
	36	696.2	72.4	836	13	BI764715	BI764715	603051051
	37	694.8	72.2	858	13	BI764309	BI764309	603046210
	38	694.2	72.2	933	14	BQ711272	BQ711272	AGENCOURT
	39	691.6	71.9	923	14	BQ709196	BQ709196	AGENCOURT
	40	690.4	71.8	852	12	BG745626	BG745626	602723873
	41	690.2	71.7	894	12	BG746247	BG746247	602723777
	42	690	71.7	810	9	AL545157	AL545157	AL545157
	43	689.2	71.6	914	13	BM007620	BM007620	603617037
	44	688.6	71.6	862	12	BG758749	BG758749	602713108
	45	687.6	71.5	844	12	BG685433	BG685433	602637288

ALIGNMENTS

RESULT 1  
AL554570  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AL554570 962 bp mRNA linear EST 16-FEB-2001  
AL554570 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DI083YJ08 5  
prime, mRNA sequence.

AL554570  
AL554570 GI:12895480  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 962)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

1. .962  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DI083YJ08"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="vector: pCMVSPORT 6; site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end



Db	241	CAACAGGGCCACTGGCATCCAGACAGGTTCAAGTGGTGGTCTGGGACAGACTTCAC	300	/organism="Homo sapiens"
QY	329	TCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCATTATTAATCTGTCAGCAATATTT	388	/db_xref="taxon:9606"
Db	301	TCTCACCATCAGCAGACTGGAGCCTGAAGACTTTGCAATGTAATCTGTCAGCAGTGA	360	/clone="IMAGE:6279586"
QY	389	TACTACTCCGTACACTTTTGGCCAGGGGACAGGCTGGAGATCAACGAAGTGTGGCTGC	448	/clone_lib="NIH_MGC_113"
Db	361	TACTCGCTTTTCACTTTCGGCCCTGGGACCAAGGTGGATATCAACGAAGTGTGGCTGC	420	/lab_host="DH10B (phage-resistant)"
QY	449	ACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCGCCTCTGT	508	/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
Db	421	ACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCGCCTCTGT	480	EcoRI; cDNA made by oligo-dT priming. Directionally cloned
QY	509	TGTGTGCCTGCTGAATTAATCTTATCCAGAGAGGCCAAAGTACAGTGGAAAGTGGATAA	568	into EcoRI/XhoI sites using the following 5' adaptor:
Db	481	TGTGTGCCTGCTGAATTAATCTTATCCAGAGAGGCCAAAGTACAGTGGAAAGTGGATAA	540	GGCAGGAG(G). Library constructed by Ling Hong in the
QY	569	CGCCCTCCAATCGGTTAATCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC	628	laboratory of Gerald M. Rubin (University of California,
Db	541	CGCCCTCCAATCGGTTAATCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC	600	Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
QY	629	CTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTA	688	Superscript II RT (Life Technologies). Note: this is a
Db	601	CTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTA	660	NIH_MGC Library."
QY	689	CGCCTGCGAAGTCAACCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGG	748	BASE COUNT
Db	661	CGCCTGCGAAGTCAACCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGG	720	221 a 277 c 221 g 204 t
QY	749	AGAGTGTAGAGGGAGAAAGTGCCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCA	808	ORIGIN
Db	721	AGAGTGTAGAGGGAGAAAGTGCCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCA	780	Query Match 81.9%; Score 787.6; DB 14; Length 923;
QY	809	TCCTTTGGCCTCTGACCCCTTTTTCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTC	868	Best Local Similarity 94.3%; Pred. No. 3e-188;
Db	781	TCCTTTGGCCTCTGACCCCTTTTTC-CAGGGGAGACTACCCCTATTGCGGGCCTCCAGCTC	839	Matches 873; Conservative 0; Mismatches 44; Indels 9; Gaps 5;
QY	869	ATCTTTACCTCAGCCCTCCCTCCTCTCCCTTAAATATGCTAATGTTGGAGGAG	926	
Db	840	ATCTTTACCTCAGCCCTCCCTCCTCTCCCT-TGGGGTTAATTATGCTAAAGGTGGAGGAG	896	
RESULT 3	BQ709678	LOCUS		
DEFINITION	AGENCOURT_8353494 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6279586			
ACCESSION	BQ709678			
VERSION	BQ709678.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	1 (bases 1 to 923)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: Dr. Mark Watson cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LCM2469 row: g column: 11 High quality sequence stop: 718.			
FEATURES	location/Qualifiers			
source	1. .923			









Db 71 ACTCTGGCTCCAGATACCACCGGAGAAATTGTGTGACGAGCTCTCCAGGCCACCTGTGC 130  
QY 149 TTTGTCTCCAGGGGAAAGAGCCACCTCTCTCTGTCAGGGCCAGTCAAGTGTTAGCAGC-- 206  
Db 131 TTTGTCTCCAGGGGAAAGAGCCACCTCTCTCTGTCAGGGCCAGTCAAGTGTAGCAGC 190  
QY 207 -TACTTAGCTGGTACCAACAGAACTGGCCAGGCTCCAGGCTCCTCATCTATGATGC 265  
Db 191 TTACTTAGCTGGTATCAGCAGAAAGCTGGCCAGGCTCCCGGCTCCTCATCCATGGTGC 250  
QY 266 ATCCAACAGGGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGGACAGACTT 325  
Db 251 ATCCAGCAGGGCCACTGGCATCCACAGACAGGTTTCAGTGGCAGTGGTCTGGGACAGACTT 310  
QY 326 CACTCTCACCATCAGCAGACTGGAGCCCGGAGGATGTGGCAGTTTATTACTGTCAAGATA 385  
Db 311 CACTCTCACCATCAGCAGACTGGAGCCCTGAAGATTTTTCAGTGTATGTCTGTCAAGATA 370  
QY 386 TTTTACTACTCCGTACACTTTTGGCCAGGGGACCGGAGGATGATGATGATGATGATGATG 445  
Db 371 TGATACCTCACAATTCACTTTCCGCCCTGGGACCAAGGTGGATATGAAACGAACTGTGGC 430  
QY 446 TGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTC 505  
Db 431 TGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTC 490  
QY 506 TGTGTGTGCTGTGTAATAACTTCTATCCAGAGAGGCAAGTACAGTGGGAAAGTGGGA 565  
Db 491 TGTGTGTGCTGTGTAATAACTTCTATCCAGAGAGGCCANAGTACAGTGGGAAAGTGGGA 550  
QY 566 TAACGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAG 625  
Db 551 TAACGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAG 610  
QY 626 CACTACAGCCTCAGCAGCAGCCTGACGCTGAGCAAGGAGCAGTACGAGAAACACAAAGT 685  
Db 611 CACTACAGCCTCAGCAGCAGCCTGACGCTGAGCAAGGAGCAGTACGAGAAACACAAAGT 670  
QY 686 CTACGCCCTGCGAAGTCAACCCATCAGGGCCTGAGCTGCGCCGTCACAAAGAGCTTCAACAG 745  
Db 671 CTACGCCCTGCGAAGTCAACCCATCAGGGCCTGAGCTGCGCCGTCACAAAGAGCTTCAACAG 730  
QY 746 GGGAGAGTGTAGAGGGAGAAAGTGCCCCACCTGCTCCTCAGTTCAGGCTGACCCCTC 805  
Db 731 GGGAGAGTGTAAAGGGAGAAAGTGCCCCACCTGCTCCTCAGTTCAGGCTGACCCCTC 790  
QY 806 CCATCCTTTGGCCTCTGACCCCTTTTCCAGAGGGGACCTACCCCTATTGCGGTCTCCAG 865  
Db 791 CCATCCTTTGGCCTCTGACCCCTTTTCCAGAGGGGACCTACCCCTATTGCGGTCTCCAG 848  
QY 866 CTCATCTTTACCTCAGCCCTCCTCCTCCTTGGCTTTAATATGCTAATGTTGGAGGA 925  
Db 849 CTCAT-TTTAACCTCAGCCCTCCTCCTCCTTCTCCT--GGNTTTAATATGCTAAGGTGGAGGA 905  
QY 926 GAATGAATAAATAAAGTGAATCTTTGC 952  
Db 906 GAATGGATAAATAAAGTGAATCTTTGC 932

RESULT 7  
BO712054  
LOCUS  
DEFINITION BO712054 926 bp mRNA linear EST 16-JUL-2002  
AGENCOURT\_8353009 NIH\_MGC\_113 Homo sapiens cdna clone IMAGE:6277900  
5', mRNA sequence.  
ACCESSION BO712054  
VERSION BO712054.1 GI:21850953  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 926)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Mark Watson  
cdna Library Preparation: Rubin Laboratory  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM2465 row: a column: 05  
High quality sequence stop: 671.

FEATURES  
source

Location/Qualifiers  
1. .926  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6277900"  
/lab\_host="NIH\_MGC\_113"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: spleen; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cdna made by oligo-dt priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

BASE COUNT 218 a 273 c 221 g 209 t 5 others  
ORIGIN

Query Match 79.8%; Score 768; DB 14; Length 926;  
Best Local Similarity 92.8%; Pred. No. 2.6e-183;  
Matches 849; Conservative 0; Mismatches 55; Indels 11; Gaps 4;

QY 35 TTAGACCCAGAGGGAAACCATGGAAGCCCCAGCTCAGCTTCTCTCTCCTCCTGCTACTCTG 94  
Db 1 TTAGACCCAGAGGGAAACCATGGAAGCCCCAGCTTCTCTCTCCTCCTGCTACTCTG 60  
QY 95 GCTCCAG-----ATACCACGGGAGAAATTGTGTGACACAGTCTCCAGCCACCTGTC 148  
Db 61 GCTCCAGTCTCAGATACACCGGAGAAATTGTGTGACGAGTCTCCAGGACCTGTC 120  
QY 149 TTTGTCTCCAGGGGAAAGAGCCACCTCTCCTGAGGGCCAGTCAGAGTGTAA--GCAG 205  
Db 121 TTTGTCTCCAGGGGAAAGAGCCACCTCTCCTGAGGGCCAGTCAGATGTAGCCGCG 180  
QY 206 CTACTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGC 265  
Db 181 CTCCTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGC 240  
QY 266 ATCCAACAGGGCCACTGGCATCCCAACCCAGGTTCAAGTGGCAGTGGTCTGGGACAGACTT 325  
Db 241 ATTCAGTAGGGCCACTGGCATCCCAACAGGTTCAAGTGGTGGTCTGGGACAGACTT 300  
QY 326 CACTCTCACCATCAGCAGACTGGAGCCCGGAGGATGTGGCAGCTTTTACTGTCAAGCAATA 385  
Db 301 CACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGAGTGTATTACTGTCAAGCAATA 360  
QY 386 TTTTACTACTCCGTACACTTTTGGCCAGGGGACCGGCTGGAGATCAAAACGAACGTGGC 445  
Db 361 TGGTAGCTCACAATTTCATTTTCGGCCCTGGGACCAAGTGGATATCAAAACGAACGTGGC 420  
QY 446 TGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTC 505  
Db 421 TGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTC 480  
QY 506 TGTGTGTGCTGTGCTGAATAACTTCTATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 565  
Db 481 TGTGTGTGCTGTGCTGAATAACTTCTATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
QY 566 TAACGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAG 625  
|||||

















```
source
1. .913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6214481"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      214 a      281 c      213 g      205 t
ORIGIN

Query Match
Best Local Similarity 76.8%; Score 738.4; DB 14; Length 913;
Matches 838; Conservative 0; Mismatches 66; Indels 10; Gaps 5;

QY 35 TTAGGACCCAGAGGGAACCATGGAAGCCCGAGCTCAGCTTCTCTTCTCCTGCTACTCTG 94
Db 1 TTAGGACCCAGAGGG-TCCATGGAACCCCGAGCGAGCTTCTCTTCTCCTGCTACTCTG 59

QY 95 GCTCCAGATACACCGGAGAAATTGTTTGACACAGTCTCCAGCCACCCCTGTCTTTGTC 154
Db 60 GCTCCAGATACACCGGAGAAATTGTTTGACGAGTCTCCAGGACCCCTGTCTTTGTC 119

QY 155 TCCAGGGGAAAGAGCCACCTCTCCTGCGAGGCGAGTCAGAGTGT---AGCAGCTACTT 211
Db 120 TCCAGGGGAAAGAGCCACCTCTCCTGCGAGGCGAGTCAGAGTGTAGCAGCAGCTACTT 179

QY 212 AGCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCAA 271
Db 180 AGCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCAG 239

QY 272 CAGGGCCACTGGCATCCACCCAGGTTTCACTGGCAGTGGTCTGGGACAGACTTCACTCT 331
Db 240 CAGGGCCACTGGCATCCACAGACAGGTTCACTGGTGTAGTGGTCTGGGACAGACTTCACTCT 299

QY 332 CACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTTACTGTGTCAGCAATATTTAC 391
Db 300 CACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTAATGTCAGCAATATGTTAC 359

QY 392 TACTCCGTACACTTTTGGCCAGGGGACCAGGCTGGAGATCAACAGAACTGTGGTGCACC 451
Db 360 CTCCTATTCACTTTCGGCCCTGGGACCAAACTGGATATCAATGAACTGTGGTGCACC 419

QY 452 ATCTGTCTTCACTTCCCGCCATCTGATGAGCAGTGGAAATCTGGAATCTGCTCTGTGT 511
Db 420 ATCTGTCTTCACTTCCCGCCATCTGATGAGCAGTGGAAATCTGGAATCTGCTCTGTGT 479

QY 512 GTGCTGTGTAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGTGGATAACGC 571
Db 480 GTGCTGTGTAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGTGGATAACGC 539

QY 572 CCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTA 631
Db 540 CCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTA 599

QY 632 CAGCCTCAGCAGCACCCCTGACGCTGAGCAAGCAGACTACGAGAAACAAAGTCTACGC 691
Db 600 CAGCCTCAGCAGCACCCCTGACGCTGAGCAAGCAGACTACGAGAAACAAAGTCTACGC 659

QY 692 CTGCGAAGTCAACCATCAGGCGCTGAGCTCGCCGCTCACAAAGAGCTTCAACAGGGGAGA 751
Db 660 CTGCGAAGTCAACCATCAGGCGCTGAGCTCGCCGCTCACAAAGAGCTTCAACAGGGGAGA 719

QY 752 GTGTTAGAGGGAGAGTGGCCCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCC 811
Db 720 GTGTTAAAGGGAGAGTGGCCCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCC 779
```

```
QY 812 TTGGCCTCTGACCC-TTTTTCACAGGGGACCTACCCCTATT--GGGTCTCTCCAGCTC 868
Db 780 TTGGCCTCTGACCCTTTTCACAGGGGACCTACCCCTATTTTGGGTCTCTCCAGCTC 839

QY 869 ATCTTTACCT---CACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 925
Db 840 ATCTTTACCTTAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899

QY 926 GAATGAATAAATAA 939
Db 900 GAAGGAAGAAATGA 913

RESULT 15
BG754043
LOCUS
DEFINITION
602709571F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4846265 5',
mRNA sequence.
ACCESSION
BG754043
VERSION
BG754043.1 GI:14064696
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 869)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1686 row: m column: 18
High quality sequence stop: 853.
FEATURES
Location/Qualifiers
1..869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4846265"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      200 a      269 c      208 g      192 t
ORIGIN
```

```
Query Match
Best Local Similarity 76.1%; Score 732; DB 12; Length 869;
Matches 820; Conservative 0; Mismatches 40; Indels 14; Gaps 5;

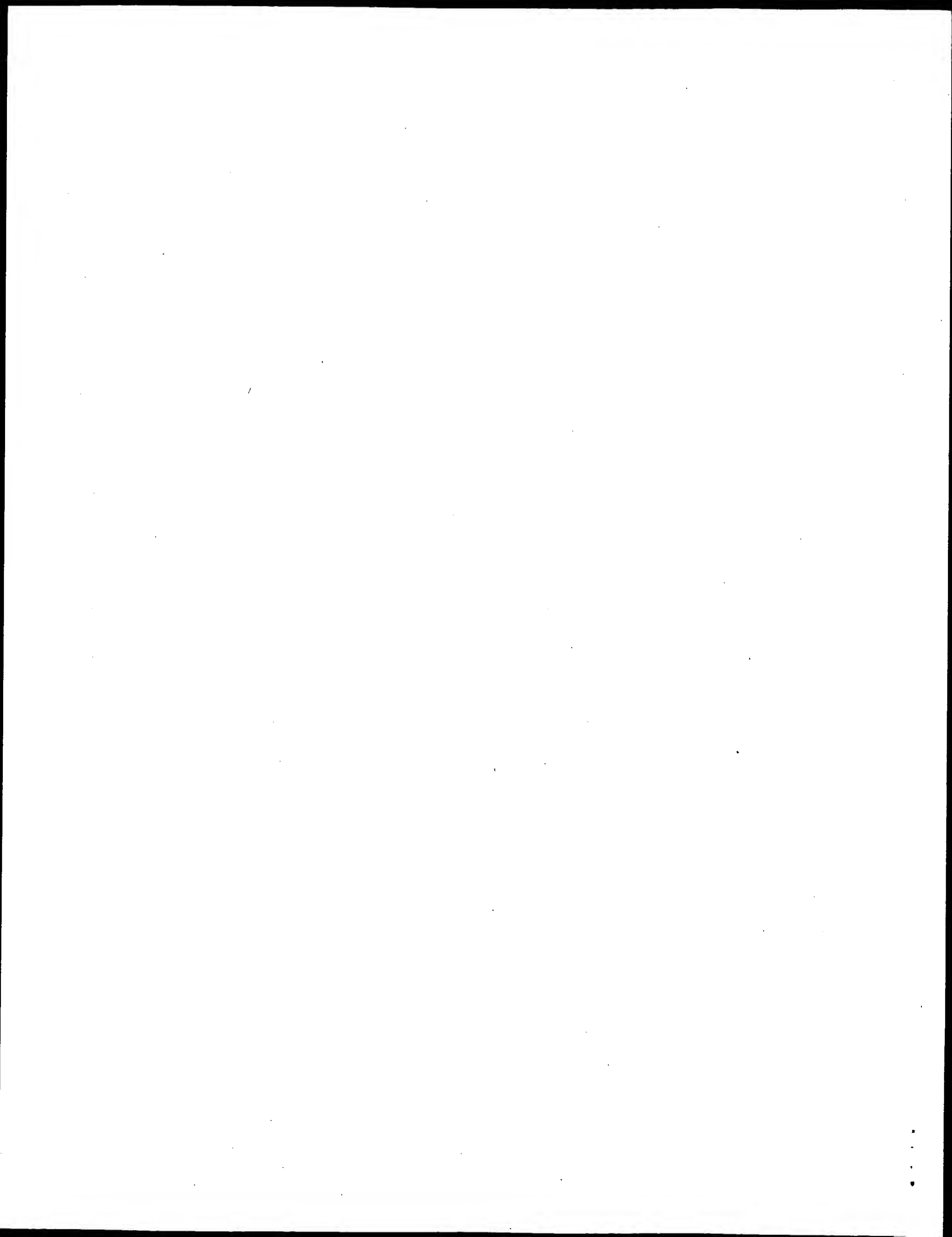
QY 56 GGAAGCCCCAGCTCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 115
Db 1 GGAAGCCCCAGCGCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 60

QY 116 AATTGTGTGACACAGTCTCCAGCCACCTGTCTTTGTCTCTCCAGGGGAAAGAGCCACCT 175
Db 61 AATTGTGTGACGAGTCTCCAGGACCCCTGTCTTTGTCTCTCCAGGGGAAAGAGCCACCT 120

QY 176 CTCCTGCAGGGCCAGTCAGAGTGT---AGCAGCTACTAGCCTGGTACCAACAGAAACC 232
```



Db 121 CTCCTGCAGGGCCAGTCAGAGTGTAGCAGCAGCTACTTAGCCTGGTACCCAGCAGAAACC 180  
QY 233 TGGCCAGGCTCCAGGCTCCTCATCTATGATGTCATCCAAAGGGCCACTGGCATCCACC 292  
Db 181 TGGCCAGGCTCCAGGCTCCTCATCTATGATGTCATCCAAAGGGCCACTGGCATCCACA 240  
QY 293 CAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACCTCTCACCATCAGCAGACTGGAGCC 352  
Db 241 CAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACCTCTCACCATCAGCAGACTGGAGCC 300  
QY 353 CGAAGATGTGGCAGTTCATTACTGTGTCAGCAATATTTTACTACTCCGTACACTTTTGGCCA 412  
Db 301 TGAAGATTTGCACTGTATTACTGTGTCAGCAGTATGGTAGCTCACCCGTACACTTTTGGCCA 360  
QY 413 GGGGACCAGGCTGGAGATCAAAAGCAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCC 472  
Db 361 -GGGACCAAGCTGGAGATCAAAAGCAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCC 419  
QY 473 ATCTGATGAGCAGTGAATCTGGAACCTGCCTCTGTGTGTGCTGCTGAATAACTTCTA 532  
Db 420 ATCTGATGAGCAGTGAATCTGGAACCTGCCTCTGTGTGTGCTGCTGAATAACTTCTA 479  
QY 533 TCCAGAGAGGCCAAAGTACAGTGGAGGTGGAATACGCCCTCCAATCGGTAACCTCCA 592  
Db 480 TCCAGAGAGGCCAAAGTACAGTGGAGGTGGAATACGCCCTCCAATCGGTAACCTCCA 539  
QY 593 GGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAGCTACAGCCTCAGCAGCAGCCTGAC 652  
Db 540 GGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAGCTACAGCCTCAGCAGCAGCCTGAC 599  
QY 653 GCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCAACCATCAGGG 712  
Db 600 GCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCAACCATCAGGG 659  
QY 713 CCTGAGCTGCCCCGTCAAAAGAGCTTCAACAGGGGAGAGTGTAGAGGG---AGAAGT 768  
Db 660 CCTGAGCTGCCCCGTCAAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGACCAAGGTGC 719  
QY 769 GCCCCACCTGCTCCTCAGTT--CCAGCCTGACCCCTCCCATCCTTTGGCCTCTGACCC 826  
Db 720 CCCCCAATCTGCTCCTCAGTTTCCAGCCTTGACCCCTTCCAATCCTTTGGCCTCTGACCC 779  
QY 827 TTTTTCACAGGGGACCTACCCCTATTGGGTCTCCAGCTCATCTTTCACTCAGCCCCC 886  
Db 780 TTTTTCACAGGGGACCTAGCCCTATTGGGGCCCTCCAGCTCATCTTTCACTCAGCCCCC 839  
QY 887 CTCCTCCTCCTTGGCTTTAATATGCTAATGTTG 920  
Db 840 CCTCTCC---TTGGTTTAAATATGCTAATGTTG 869



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 25, 2003, 15:39:40 ; Search time 212.803 Seconds  
(without alignments)  
9326.039 Million cell updates/sec

Title: US-09-806-276A-3  
Perfect score: 962  
Sequence: 1 tcgagccgattggctcgag.....gaattttgcaaaaaaaaaa 962

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues 2879534  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq1:  
11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq3:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	850.4	88.4	1184	15	US-10-158-646-60
2	845.4	87.9	970	11	US-09-859-053-37
3	830.2	86.3	948	11	US-09-859-053-33
4	816.4	84.9	1202	15	US-10-158-646-57
5	808.4	84.0	913	11	US-09-822-830A-531
6	792.4	82.4	1244	11	US-09-954-456-771
7	767.2	79.8	1001	15	US-10-198-846-14013
8	746.8	77.6	974	11	US-09-859-053-29
9	746	77.5	1775	15	US-10-158-646-64
10	738.8	76.8	1033	10	US-09-799-514-2
11	726.4	75.5	1230	15	US-10-158-646-59
12	724.6	75.3	990	10	US-09-800-729-79
13	716.8	74.5	1458	15	US-10-158-646-66
14	713.8	74.2	961	15	US-10-198-846-13125
15	713.2	74.1	941	10	US-09-800-729-81
16	710.4	73.8	928	15	US-10-221-945-5

17	708.6	73.7	829	15	US-10-158-646-4	Sequence 4, Appli
18	688.2	71.5	853	15	US-10-158-646-68	Sequence 68, Appl
19	688	71.5	1045	15	US-10-198-846-13629	Sequence 13629, A
20	684.4	71.1	968	12	US-09-992-600A-7	Sequence 7, Appli
21	684.4	71.1	968	12	US-09-924-340-7	Sequence 7, Appli
22	684.4	71.1	968	15	US-10-000-489-7	Sequence 7, Appli
23	684.4	71.1	968	15	US-10-000-986-7	Sequence 7, Appli
24	681.8	70.9	1775	15	US-10-158-646-64	Sequence 64, Appl
25	680	70.7	2272	15	US-10-158-646-63	Sequence 63, Appl
26	658.8	68.5	2272	15	US-10-158-646-63	Sequence 63, Appl
27	652.2	67.8	705	15	US-10-153-382-14	Sequence 14, Appl
28	650.2	67.6	743	15	US-10-198-846-13869	Sequence 13869, A
29	640.8	66.6	708	15	US-10-153-382-6	Sequence 6, Appli
30	635	66.0	819	15	US-10-158-646-65	Sequence 65, Appl
31	633.2	65.8	962	15	US-10-198-846-10845	Sequence 10845, A
32	629.8	65.5	702	15	US-10-153-382-10	Sequence 10, Appl
33	618.2	64.3	944	15	US-10-198-846-13859	Sequence 13859, A
34	597.6	62.1	672	12	US-09-453-234-83	Sequence 83, Appl
35	596	62.0	672	12	US-09-453-234-81	Sequence 81, Appl
36	596	62.0	672	12	US-09-453-234-87	Sequence 87, Appl
37	596	62.0	672	12	US-09-453-234-89	Sequence 89, Appl
38	594.4	61.8	672	12	US-09-453-234-35	Sequence 35, Appl
39	594.4	61.8	672	12	US-09-453-234-45	Sequence 45, Appl
40	593.8	61.7	584	15	US-10-198-846-13823	Sequence 13823, A
41	587.6	61.1	678	12	US-09-453-234-85	Sequence 85, Appl
42	586	60.9	678	12	US-09-453-234-49	Sequence 49, Appl
43	579.6	60.2	678	12	US-09-453-234-73	Sequence 73, Appl
44	578	60.1	678	12	US-09-453-234-79	Sequence 79, Appl
45	574.6	59.7	645	12	US-09-972-656-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1  
US-10-158-646-60  
; Sequence 60, Application US/10158646  
; Publication No. US20030073105A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy K.W.  
; APPLICANT: Sornasse, Thierry  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0030-1 US  
; CURRENT APPLICATION NUMBER: US/10/158,646  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: 60/295,239  
; PRIOR FILING DATE: 2001-05-31  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PERL Program  
; SEQ ID NO 60  
; LENGTH: 1184  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20030073105A1 1135037.21  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 44, 65  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-158-646-60

Query Match	88.4%;	Score	850.4;	DB	15;	Length	1184;
Best Local Similarity	96.2%;	Pred. No.	1.2e-238;				
Matches	893;	Conservative	0;	Mismatches	31;	Indels	4;
						Gaps	2;
QY	29	GCTCAGTTAGGACCCAGAGGGAACCATGGAAGCCCGAGCTCAGCTTCCTCTCCTGCT	88				
Db	257	GCTCAGTTAGGACCCAGAGGGAACCATGGAAGCCCGAGCTCCTCTCCTGCT	316				
QY	89	ACTCTGGCTCCAGATACCCAGGAGAAATTTGTTGACACAGTC-TCCAGCCACCTGT	147				
Db	317	ACTCTGGCTCCAGATACCCAGGAGAAATTTGTTGACACAGTC-TCCAGCCACCTGT	376				







Query Match	84.9%;	Score 816.4;	DB 15;	Length 1202;
Best Local Similarity	95.7%;	Pred. No. 1.1e-228;		
Matches 907;	Conservative 0;	Mismatches 26;	Indels 15;	Gaps 6;
QY 29	GCTCAGTTAGGACCCAGAGGGAACCAT-GGAAGCCCCAGCTCAGCTTCTCTTCTCCTCCTGC	87		
Db				
256	GCTCAGTTAGGACCCAGAGGGAACCATGGGAAACCCAGCGAGCTTCTCTTCTCCTCCTGC	315		
QY	88 TACTCTGGCTCC-----CAGATACCACCGGAGAAAATTGTGTGACACAGTC-TCCAGCC	140		
Db				
316	TACTCTGGCTCCAGTCGCAGTTATACCGGAGAAAATTGTGTGACGAGCTTCCAGGC	375		
QY	141 ACCCTGCTTTGTCTCCAGGGGAAAGAGCCACCCCTCTCTGACGGGCCAGTCAGAGTGT	200		
Db				
376	ACCCTGCTTTGTCTCCAGGGGAAAGAGCCACCCCTCTCTGACGGGCCAGTCAGAGTGT	435		
QY	201 ---AGCAGCTACTTAGCCTGGTACCAACAGAAAACCTGGCCAGGCTCCCAGGCTCTCATC	257		
Db				
436	AGCAGAGCTACTTAGCCTGGTACCAACAGAAAACCTGGCCAGGCTCCCAGGCTCTCATC	495		
QY	258 TATGATGCATCCAACAGGGGCACCTGGCATCCCACCCAGGTTTCAGTGGCAGTCTGGG	317		
Db				
496	TATGGTGCATCCAGCAGGGCCACTGGCATCCCACAGAGTTTCAGTGGCAGTCTGGG	555		
QY	318 ACAGACTTCACCTCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTATTACTGT	377		
Db				
556	ACAGACTTCACCTCTCACCATCAGCAGCCCTGCAGCTCTGAAGATTTTGCAGTGTATTACTGT	615		
QY	378 CAGCAATATTTTA---CTACTCCGTACACTTTTGGCCAGGGACCCAGGCTGGAGATCAAA	434		
Db				
616	CAGCAGTATAATAACTGGCCCTCCGTACACTTTTGGCCAGGGACCAAGCTGGAGATCAAA	675		
QY	435 CGAACTGTGGCTGCACCATCTGTCTTCACTCTCCGCCCATCTGATGAGCAGTTGAAATCT	494		
Db				
676	CGAACTGTGGCTGCACCATCTGTCTTCACTCTCCGCCCATCTGATGAGCAGTTGAAATCT	735		
QY	495 GGAAGTGCCTCTGTGTGTGCCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAG	554		
Db				
736	GGAAGTGCCTCTGTGTGTGCCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAG	795		
QY	555 TGGAAAGTGGATAACGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGAC	614		
Db				
796	TGGAAAGTGGATAACGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGAC	855		
QY	615 AGCAAGGACAGCACCTACAGCCCTCAGCAGCACCTTGACGCTGAGCAAAAGCAGACTACGAG	674		
Db				
856	AGCAAGGACAGCACCTACAGCCCTCAGCAGCACCTTGACGCTGAGCAAAAGCAGACTACGAG	915		
QY	675 AAACACAAAGTCTACGCCCTGCGAAGTCAACCATCAGGGGCTGAGCTCGCCCGTCACAAAG	734		
Db				
916	AAACACAAAGTCTACGCCCTGCGAAGTCAACCATCAGGGGCTGAGCTCGCCCGTCACAAAG	975		
QY	735 AGCTTCAACAGGGGAGAGTGTAGAGGGAGAGTGCCCCACCTGCTCCTCAGTTCACAGC	794		
Db				
976	AGCTTCAACAGGGGAGAGTGTAGAGGGAGAGTGCCCCACCTGCTCCTCAGTTCACAGC	1035		
QY	795 CTGACCCCTCCCATCTTTGGCCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTG	854		
Db				
1036	CTGACCCCTCCCATCTTTGGCCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTG	1095		
QY	855 CGGTCTCCAGCTCATCTTTCACCTCACCCCTCCTCCTCCTTGGCTTAAATATGCTA	914		
Db				
1096	CGGTCTCCAGCTCATCTTTCACCTCACCCCTCCTCCTCCTTGGCTTAAATATGCTA	1155		
QY	915 ATGTTGGAGGAGAAATGAATAATAAAGTGAATCTTTGCAAAAAAAAAA	962		
Db				
1156	ATGTTGGAGGAGAAATGAATAAAT-AAGTGAATCTTTGCAAAAAAAAAA	1202		

RESULT 5  
US-09-822-830A-531  
; Sequence 531, Application US/09822830A

```

; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 531
; LENGTH: 913
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-531

Query Match          84.0%; Score 808.4; DB 11; Length 913;
Best Local Similarity 94.0%; Pred.No. 2.2e-226;

```

Query Match	84.0%;	Score 808.4;	DB 11;	Length 913;
Best Local Similarity	94.0%;	Pred. No. 2.2e-226;		
Matches 852;	Conservative	0;	Mismatches 51;	Indels 3; Gaps 1;
QY	50	AACCATGGAAGCCCCAGCTCAGCTTCTCTTCTCTCTCTGCTACTCTGGCTCCAGATACCAC	109	
Db	1			
QY	110	CGGAGAAATTGTGTGACACAGTCTCCAGGCCACCCCTGTCTTTGTCTCCAGGGGAAAGAGC	169	
Db	61			
QY	170	CACCTCTCCTGCAGGGCCAGTCAGAG---TGTAGCAGCTACTTAGCCTGGTACCAACA	226	
Db	121			
QY	227	GAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGCCACTGGGCAT	286	
Db	181			
QY	287	CCCACCCAGTTTCACTGGCAGTGGGTCTGGGACAGACTTCACCTCTCACCATCAGCAGACT	346	
Db	241			
QY	347	GGAGCCCGAAGATGFGCACTTTATTACTGTGAGCAATATTTTACTACTCCGTACACTTT	406	
Db	301			
QY	407	TGGCCAGGGACAGGCTGGAGATCAAACGAACTGTGGCTGCACCATCTGTCTTCATCTT	466	
Db	361			
QY	467	CCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTTGTGTGCTGCTGAATAA	526	
Db	421			
QY	527	CTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATAACGCCCTCCAATCGGGTAA	586	
Db	481			
QY	587	CTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCCTACAGCCTCAGCAGCAC	646	
Db	541			
QY	647	CCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCAACCCA	706	
Db	601			





```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-14013

Query Match 79.8%; Score 767.2; DB 15; Length 1001;
Best Local Similarity 91.2%; Pred. No. 2.6e-214;
Matches 861; Conservative 0; Mismatches 73; Indels 10; Gaps 4;

29 GCTCAGTTAGGACCCAGAGGGAACCATGGAAGCCCCAGCTCAGCTTCTCTTCTCCTCTGCT 88
|||||
56 GCTCAGTTAGGACCCAGAGGGAACCATGGAAGCCCCAGCTCAGCTTCTCTTCTCCTCTGCT 115
|||||
89 ACTCTGGCTCCCAGATACACACCCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCTGTC 148
|||||
116 ACTCTGGCTCCCAGATACACACCCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCTGTC 175
|||||
149 TTTGTCTCCAGGGGAAAGAGCCACCCTCTCTCTGAGGGCCAGTCAGAGTGTAGCAGCTA 208
|||||
176 TGTGTCTCCAGGGGAAAGAGCCACCCTCTCTCTGAGGGCCAGTCAGAGTGTGGCAGCAA 235
|||||
209 CTTAGCCCTGGTACCAACAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGATGCATC 268
|||||
236 CTTAGCCCTGGTATCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTTGGTGCATC 295
|||||
269 CAACAGGGCCACTGGCATCCCAACCCAGG-TTCAGTGGCAGTGGGCTG-GGACAGACTTC 326
|||||
296 CACCAGGGCCACTGGTATCCCAAGCCAGGATTCAGTGGCGGTGGTCCAGAGCAGAGTTC 355
|||||
327 ACTCTCACCATCAGCAGACTGGAGCCCGAAGATGTGSCACTTTATTACTGTCAGCAATAT 386
|||||
356 ACTCTCACCATCAGCAGTATGCAGTCTGAAGATTTTGCAATTTATTCTCTGTCAGCATAT 415
|||||
387 TTTACTACTCC-----GTACACTTTTGGCCAGGGGACAGGCTGGAGATCAAAACGAAC 439
|||||
416 AATAACTGGCCCTCCGAAGTTTCACCTTCGGCCCAAGGACCAAGGTGGAATCAAAACGAAC 475
|||||
440 TGTGGCTGCACCATCTGCTTTCATCTTCCCGCCATCTGATGAGCAGTTTGAATCTGGAAC 499
|||||
476 TGTGGCTGCACCATCTGCTTTCATCTTCCCGCCATCTGATGAGCAGTTTGAATCTGGAAC 535
|||||
500 TGCCTCTGTTGTGCTGCTGCTGAATACTTCTATCCCAAGAGAGGCCAAAGTACAGTGGAA 559
|||||
536 TGCCTCTGTTGTGCTGCTGCTGAATACTTCTATCCCAAGAGAGGCCAAAGTACAGTGGAA 595
|||||
560 GGTGGATAACGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAA 619
|||||
596 GGTGGATAACGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAA 655
|||||
620 GGACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAGCAGACTACGAGAAACA 679
|||||
656 GGGCAGCCCTCAAGCCTCAGCAGCCCTGACGCTGGGCAAGCAGACTACGAGAAACC 715
|||||
680 CAAAGTCTACGCCCTCGAAGTCAACCCATCAGGSCCTGAGCTCGCCGTCACAAAGAGCTT 739
|||||
716 CCAACTCTACGCCCTCGAAGTCAACCCATCAGGSCCTGAGCTCGCCGTCACAAAGAGCTT 775
|||||
740 CAACAGGGGAGAGTGT-AGAGGGAGAAAGTGCCTCCACCTGCTCCTCAGTTCACAGCCTGA 798
|||||
776 CAACAGGGGAGAGTGTAAAGGGGAAAGGGCCCCCCCCCTGCTCCTCAGTTCACAGCCTGA 835
|||||
799 CCCCCTCCCATCCTTTGGCCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGT 858
|||||
836 CCCCCTCCCATCCTTTGGCCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGT 895
|||||
859 CCTCCAGTTCATCTTTCACCTACCCCTCCCTCCTCCTTGGCTTTAATATGCTAATGT 918
|||||
896 CCTCCAGTTCATCTTCCCTCCCTCCCTCCTCCTTGGCTTTAATATGCTAATGT 955
|||||
919 TGGAGGAGATGAATAAATAAGTGAATCTTTGCAAAAAA 962
|||
956 TGGGGGGAATGAATAAATAAGTGAATTTTGGCAAAAAA 999
|||

```











```
QY 94 GGCTCCAGATACACCGGAGAAATGTGTGACACAGTCTCCAGGCCACCCCTGTCTTGT 153
|||||
Db 588 GGCTCCAGGCGCAAGTGTGACATGACATGACCCAGTCTCCTTCCACCCCTGTCTGCTT 647
|||||
QY 154 CTCCAGGGGAAAGAGCCACCCCTCTCTGCGAGGGCCAGT-CAGAGTGTAGCAG----- 205
|||
Db 648 CTGTCGGAGACAGAGTCACTATCACTGTGCGGGCCAGTGCAGAGTGTAGCAGCACCGGC 707
|||
QY 206 -----CTACTTAGCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTAT 260
|||
Db 708 AGTAACCTGGTGGCTGGTACCAACAGAAACCCAGGAAACCCCTAAGCTCCTGATCTAT 767
|||
QY 261 GATGCATCCAAACAGGGCCA-CTGGCATCCACCCAGGTTCAAGTGGCAGTGGGTCTGGGAC 319
|||
Db 768 AAGCGCTACTTTAGAAAGATGGGGTCCCATCAAGGTTCAAGTGGCAGTGGGTCTGGGAC 827
|||
QY 320 AGACTTCACTCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTTATTACTGTCA 379
|||||
Db 828 AGACTTCACTCTCACCATCAGCAGACTGGAGCCTGATGATTTTGCAGTGTATTACTGTCA 887
|||||
QY 380 GCAATATTTACTACTCCGTACACTTTTGGCCAGGGGACAGGCTGGAGATCAAAACGAAC 439
|||||
Db 888 GCAGTATAATAACATGCCTCGGACGTTGCGCCCTGGGACCAAGCTGGAGATCAAAACGAAC 947
|||||
QY 440 TGTGGCTGCACCATCTGTCTTCTATCTTCCCGCCATCTGTGATGAGCAGTTGAAATCTGGAAC 499
|||||
Db 948 TGTGGCTGCACCATCTGTCTTCTATCTTCCCGCCATCTGTGATGAGCAGTTGAAATCTGGAAC 1007
|||||
QY 500 TGCCTCTGTGTGCTGCTGCTGAATAACTTCTATCCACAGAGAGGCCAAAGTACAGTGGAA 559
|||||
Db 1008 TGCCTCTGTGTGCTGCTGCTGAATAACTTCTATCCACAGAGAGGCCAAAGTACAGTGGAA 1067
|||||
QY 560 GGTGGATAACGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAA 619
|||||
Db 1068 GGTGGATAACGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAA 1127
|||||
QY 620 GGACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACA 679
|||||
Db 1128 GGACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACA 1187
|||||
QY 680 CAAAGTCTACGCTGCGAAGTCAACCCATCAGGGCCCTGAGCTCGCCCGTCAACAAAGAGCTT 739
|||||
Db 1188 CAAAGTCTACGCTGCGAAGTCAACCCATCAGGGCCCTGAGCTCGCCCGTCAACAAAGAGCTT 1247
|||||
QY 740 CAACAGGGGAGAGTGTAGAGGGAGAAAGTGCCCGCCACCTGCTCCTCAGTTCACGCTGAC 799
|||||
Db 1248 CAACAGGGGAGAGTGTAGAGGGAGAAAGTGCCCGCCACCTGCTCCTCAGTTCACGCTGAC 1307
|||||
QY 800 CCCCTCCCATCTTTGGCCCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTC 859
|||||
Db 1308 CCCCTCCCATCTTTGGCCCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTC 1367
|||||
QY 860 CTCAGCTCATCTTTCACTCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 919
|||||
Db 1368 CTCAGCTCATCTTTCACTCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1427
|||||
QY 920 GGAGGAGAAATGAATAAATAAGTGAATCTT 949
|||||
Db 1428 GGAGGAGAAATGAATAAATAAGTGAATCTT 1457
|||||
```

RESULT 14  
US-10-198-846-13125  
; Sequence 13125, Application US/10198846  
; Publication No. US200309974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13125  
; LENGTH: 961  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-198-846-13125

Query Match 74.2%; Score 713.8; DB 15; Length 961;  
Best Local Similarity 90.4%; Pred. No. 1.1e-198;  
Matches 856; Conservative 0; Mismatches 72; Indels 19; Gaps 8;

```
QY 24 CTCGAGCTCAGTTAGGACCCAGAGGGAAACCATGGAAGCCCAAGCTCAGCTTCTCTTCTC 83
|||
Db 11 CACGCGTCCGGTTAGGACCCAGAGGGAAACCATGGAAGCCCAAGCTTCTCTTCTC 70
|||
QY 84 CTGCTACTCTGGCTCCAGATACCACCGGAGAAATGTGTTGACACAGTCTCCAGCCACC 143
|||||
Db 71 CTGCTACTCTGGCTCCAGAGACCCAGCGGAGAAATGTGTTGACGAGTCTCCAGGCACC 130
|||||
QY 144 CTGCTCTTGTCTCCAGGGGAAAGAGCCACCCCTCTCTCTGAGGGCCAGTCAAGTGT 200
|||||
Db 131 CTGCTCTTGTCTCCAGGGGAAAGAGCCACCCCTCTCTCTGAGGGCCAGTCAAGTGT 190
|||||
QY 201 AGCAGCTACTTAGCTTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTAT 260
|||||
Db 191 AGCAGCTACTTAGCTTGGTATCAGCAGAAACCTGGCCAGGCTCCAGGCTCCTCATGTAT 250
|||||
QY 261 GATGCATCCAAACAGGGCCACTGGCATCCCAAGGTTCAAGTGGCAGTGGGTCTGGGACA 320
|||||
Db 251 GGTGCATCCAGCAGGGCCACTGGCATCCCAAGATTCAGTGGCAGTGGGTCTGGGACA 310
|||||
QY 321 GACTTCACTCTCACCATCAGCAGACTGGAGCCCGAAGATGTGCACTTTATTACTGTGAC 380
|||||
Db 311 GACTTCACTCTCACCATCAACAGACTGGAGCCTGATGATTTTGCAGTATATTACTGTGAC 370
|||||
QY 381 CAATATTTTACTACTCCGTACACTTTTGGCCAGGGACCAAGCTGGAGATCAAAACGAAC 440
|||||
Db 371 CACTATGCTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 430
|||||
QY 441 GTGGCTGCACCATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 500
|||||
Db 431 GTGGCTGCACCATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 490
|||||
QY 501 GCCTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 560
|||||
Db 491 GCCTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
|||||
QY 561 GTGGATAACGCCCTCCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAG 620
|||||
Db 551 GTGGATAACGCCCTCCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAG 610
|||||
QY 621 GACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACAC 680
|||||
Db 611 GACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAAGCAGACTACGAGAAACAC 670
|||||
QY 681 AAAGTCTAGCCCTCGGAAG-TCACCCATCAGGGCC-----TGAGCTCGCCCGTCACAAAGA 735
|||||
Db 671 AAAGTCTAGCCCTCGGAAGTTCACCCATCAGGGGCTCGAGGTTGCGCCGTTTCACAAAGA 730
|||||
QY 736 GCTTCAACAGGGGAGAGTGTAGAGGGAG-----AAGTCCCCCACCTGCTCTCAGT 788
|||||
Db 731 GCTTCAACAGGGGAGAGTGTAGAGGGAGGAGGAGTGTCCCCCACCTGCTCTCAGT 790
|||||
QY 789 TCCAGGCTGACCCCTCTCCCATCTT-GGCCTCTGACCCCTTTTCCACA-GGGGACCTAC 846
|||||
Db 791 TCCAGGCTGACCCCTCTCCCATCTTGGGCTCTGACCCCTTTTCCACAGGGGACCTAC 850
|||||
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 15:25:15 ; Search time 46.6192 Seconds  
(without alignments)  
6328.365 Million cell updates/sec

Title: US-09-806-276A-3  
Perfect score: 962  
Sequence: 1 tcgagccgattcgctcgag.....gaatctttgcaaaaaaaaaa 962

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	829.6	86.2	931	3	US-09-049-672A-19	Sequence 19, Appl
2	751	78.1	3819	4	US-09-042-353-393	Sequence 393, App
3	751	78.1	3819	4	US-08-758-417A-243	Sequence 243, App
4	738.8	76.8	1066	1	US-08-157-101A-4	Sequence 4, Appli
5	685	71.2	958	3	US-09-049-672A-15	Sequence 15, Appl
6	580.2	60.3	8068	4	US-09-301-593-35	Sequence 35, Appl
7	565.2	58.8	13254	1	US-08-276-852-156	Sequence 156, App
8	565.2	58.8	13254	1	US-08-276-852-170	Sequence 170, App
9	565.2	58.8	13254	1	US-08-899-575-156	Sequence 156, App
10	565.2	58.8	13254	1	US-08-899-575-170	Sequence 170, App
11	565.2	58.8	13254	1	US-08-899-575-156	Sequence 156, App
12	565.2	58.8	13254	1	US-08-899-575-170	Sequence 170, App
13	565.2	58.8	13254	5	PCT-US95-08743-156	Sequence 156, App
14	565.2	58.8	13254	5	PCT-US95-08743-170	Sequence 170, App
15	563	58.5	729	1	US-08-276-852-152	Sequence 152, App
16	563	58.5	729	1	US-08-276-852-168	Sequence 168, App
17	563	58.5	729	1	US-08-899-575-152	Sequence 152, App
18	563	58.5	729	1	US-08-899-575-168	Sequence 168, App
19	563	58.5	729	1	US-08-899-575-152	Sequence 152, App
20	563	58.5	729	1	US-08-899-575-168	Sequence 168, App
21	563	58.5	729	5	PCT-US95-08743-152	Sequence 152, App
22	563	58.5	729	5	PCT-US95-08743-168	Sequence 168, App
23	553	57.5	4691	4	US-09-591-632-43	Sequence 43, Appl
24	553	57.5	4691	4	US-09-611-451-43	Sequence 43, Appl
25	553	57.5	6166	4	US-08-591-632-51	Sequence 51, Appl
26	553	57.5	6166	4	US-09-611-451-51	Sequence 51, Appl
27	551.4	57.3	646	1	US-08-300-386A-2	Sequence 2, Appli

28	551.4	57.3	646	3	US-08-931-645-2	Sequence 2, Appli
29	551.4	57.3	646	5	PCT-US94-01258-2	Sequence 2, Appli
30	551.4	57.3	646	5	PCT-US95-11235-2	Sequence 2, Appli
31	551	57.3	642	2	US-08-480-753-5	Sequence 5, Appli
32	551	57.3	642	3	US-09-041-889-10	Sequence 10, Appl
33	551	57.3	642	3	US-08-837-058-10	Sequence 10, Appl
34	547.6	56.9	645	2	US-08-480-753-7	Sequence 7, Appli
35	545	56.7	8068	4	US-09-301-593-27	Sequence 27, Appl
36	538.6	56.0	19040	4	US-09-343-485A-3	Sequence 3, Appli
37	531.8	55.3	705	1	US-08-488-376-16	Sequence 16, Appl
38	531.8	55.3	705	2	US-08-634-223-16	Sequence 16, Appl
39	531.8	55.3	705	2	US-08-634-224-16	Sequence 16, Appl
40	531.8	55.3	705	2	US-08-634-400-16	Sequence 16, Appl
41	531.8	55.3	705	2	US-08-635-878-16	Sequence 16, Appl
42	531.8	55.3	705	2	US-08-770-057-16	Sequence 16, Appl
43	531.8	55.3	705	4	US-09-335-697B-16	Sequence 16, Appl
44	531.8	55.3	705	4	US-09-335-697B-16	Sequence 16, Appl
45	523.2	54.4	708	1	US-08-488-376-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-09-049-672A-19  
; Sequence 19, Application US/09049672A  
; Patent No. 6135941  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Yue, Henry  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/049,672A.  
; FILING DATE: HEREWITH  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cerrone, Michael C  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0497 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 931 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: COLSUCT01









QY	413	GGGGACCAAGCTGGAGATCAAACGAACCTGTGGCTGCACCATCTGTCTTCACTCTCCGGCC	472
Db	398	AGGGACCAAGGTGGAGATCAAACGAACCTGTGGCTGCACCATCTGTCTTCACTCTCCGGCC	457
QY	473	ATCTGATGAGCAGTGAATCTGGAATCGCCTCTGTGTGTGCTGCTGCTGAATAACTTCTA	532
Db	458	ATCTGATGAGCAGTGAATCTGGAATCGCCTCTGTGTGTGCTGCTGCTGAATAACTTCTA	517
QY	533	TCCAGAGAGGCCAAAGTACAGTGGAAAGGTGGATAACGCCCTCCAATCGGGTAACTCCTA	592
Db	518	TCCAGAGAGGCCAAAGTACAGTGGAAAGGTGGATAACGCCCTCCAATCGGGTAACTCCTA	577
QY	593	GGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCCTGAC	652
Db	578	GGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCCTGAC	637
QY	653	GCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTAGCCTGCGAAGTCACCCATCAGGG	712
Db	638	GCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTAGCCTGCGAAGTCACCCATCAGGG	697
QY	713	CCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTAGAGGGAGAAGTGCCC	772
Db	698	CCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTAGAGGGAGAAGTGCCC	757
QY	773	CCACCTGCTCCTCAGTTCACAGCCTGACCCCTCCCATCCTTTGGCCTCTGACCCCTTTTC	832
Db	758	CCACCTGCTCCTCAGTTCACAGCCTGACCCCTCCCATCCTTTGGCCTCTGACCCCTTTTC	817
QY	833	CACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCACCTCACCCCTCCTC	892
Db	818	CACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCACCTCACCCCTCCTC	877
QY	893	CTCCTTGGCTTTAATTATGCTAATGTTGGAGGAGAATGAATAAATAAAAGTGAATCTTGC	952
Db	878	CTCCTTGGCTTTAATTATGCTAATGTTGGAGGAGAATGAATAAATAAAAGTGAATCTTGC	937
QY	953	AAAAAAAAAA 962	
Db	938	AAAAAAAAAA 947	

RESULT 5  
 US-09-049-672A-15  
 ; Sequence 15, Application US/09049672A  
 ; Patent No. 6135941  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Yue, Henry  
 ; APPLICANT: Au-Young, Janice  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Baughn, Mariah R.  
 ; TITLE OF INVENTION: HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS  
 ; NUMBER OF SEQUENCES: 28  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/049,672A  
 ; FILING DATE: HEREWITH  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ;

[illegible]







; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,852
; FILING DATE: 18-JUL-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,302
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/954,148
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCR1452P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 156:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-276-852-156

Query Match 58.8%; Score 565.2; DB 1; Length 13254;
Best Local Similarity 86.8%; Pred. No. 6.8e-152;
Matches 634; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

QY 39 GACCCAGAGGGAACCATGGAAGCCCGAGCTCAGCTTCTTCTCCTCCTGCTACTCTGGCTC 98
Db 12522 GACACGAAGCTTACCATGGGTGTGCCACTCAGGTCCTGGGGTTGCTGCTGTGGCTT 12581

QY 99 CCAGATACCAACCGGAGAAATTTGTTGACACAGTCTCCAGCCACCTGTCTTTGTCTCCA 158
Db 12582 ACAGATGCCAGATGTGAGATCGTTCTCACGAGTCTCCAGGCACCTGTCTGTCTCCA 12641

QY 159 GGGGAAAGAGCCACCTCTCCTGCAGGGCCAGTCAGAGTGTT---AGCAGCTACTTAGCC 215
Db 12642 GGGGAAAGAGCCACCTTCTCCTGTAGGTCCAGTCACAGCATTCGAGCCGCGGTAGCC 12701

QY 216 TGGTACCAACAGAAACCTTGCCAGGCTCCAGGCTCCTCATCATGATGATGATCAACAGG 275
Db 12702 TGGTACCAACAGAAACCTTGCCAGGCTCCAGGCTCCAGGCTGGTCAATAGGTTTCCAATAG 12761

QY 276 GCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACC 335
Db 12762 GCCTCTGGCATCTCAGACAGGTTTCAGGCGCAGTGGGTCTGGGACAGACTTCACTCTCACC 12821

QY 336 ATCAGCAGACTGGAGCCCGGAAGATGTGGCACTTTATTACTGTACGAAATATTTTACTACT 395
Db 12822 ATCACCAGAGTGGAGCCCTGAAGACTTTGCACTGTACTACTGTGAGGTCTATGGTGCCTGC 12881

QY 396 CCGTACACTTTGGCCAGGGACAGGCTGGAGATCAAAACGAACTGTGGCTGCACCATCT 455
Db 12882 TCGTACACTTTGGCCAGGGACCAAACTGGAGAGGAAACGAACTGTGCCTGCACCATCT 12941

QY 456 GTCTTCACTTCCCGCCATCTGTAGCAGTGAATCTGAAATCTGGAATCTGCTGTGTGTGC 515
Db 12942 GTCTTCACTTCCCGCCATCTGTAGCAGTGAATCTGGAATCTGCTGTGTGTGC 13001

QY 516 CTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATAACGCCCTC 575
Db 13002 CTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATAACGCCCTC 13061

QY 576 CAATCGGTAACCTCCAGGAGAGTGTACAGAGCAGGACGAGCAAGGACACCTACAGC 635
Db 13062 CAATCGGTAACCTCCAGGAGAGTGTACAGAGCAGGACGAGCAAGGACACCTACAGC 13121

QY 636 CTCAGCAGCACCCCTGACCGCTGACCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGC 695
Db 13122 CTCAGCAGCACCCCTGACCGCTGACCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGC 13181

QY 696 GAAGTCACCCATCAGGCGCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGT 755
Db 13182 GAAGTCACCCATCAGGCGCTGAGATCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGT 13241

QY 756 TAGAGGGAGA 765
Db 13242 TAATTCTAGA 13251

RESULT 8
US-08-276-852-170/c
; Sequence 170, Application US/08276852
; Patent No. 5652138
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 170
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
; STREET: Mail Drop TPC8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,852
; FILING DATE: 18-JUL-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,302
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/954,148
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCR1452P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-276-852-170

Query Match 58.8%; Score 565.2; DB 1; Length 13254;
Best Local Similarity 86.8%; Pred. No. 6.8e-152;
Matches 634; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

QY 39 GACCCAGAGGGAACCATGGAAGCCCGAGCTTCTTCTCCTCCTGCTACTCTGGCTC 98
Db 733 GACACGAAGCTTACCATGGGTGTGCCACTCAGGTCCTGGGGTTGCTGCTGTGGCTT 674

QY 99 CCAGATACCAACCGGAGAAATTTGTTGACACAGTCTCCAGCCACCTGTCTTTGTCTCCA 158
Db 673 ACAGATGCCAGATGTGAGATCGTTCTCACGAGTCTCCAGGCACCTGTCTGTCTCCA 614











673	ACAGATGCCAGATGTGAGATCGTTCTCACGCAGTCTCCAGGCACCCCTGTCTCTGTCTCCA	614	
159	GGGGAAGAGCCACCCCTCTCCTCGAGGGCCAGTCAGAGTGT---	AGCAGCTACTTAGCC	215
613	GGGGAAGAGCCACCCCTCTCCTGTAGGTCCAGTCACAGCATTCGCAGCGCGCGTAGCC	554	
216	TGGTACCAACAGAAACCTGGCCAGGCTCCAGGTCCTCATCTATGATCGATCCAACAGG	275	
553	TGGTACCAGCACAAACCTGGCCAGGCTCCAAGGCTGGTCATACATGGTGTTCCAATAGG	494	
276	GCCACTGGCATCCCAACCCAGGTTCAAGTGGCAGTGGTCTGGGACAGACTTCACCTCACC	335	
493	GCCTCTGGCATCTCAGACAGGTTCAAGGCGAGTGGTCTGGGACAGACTTCACCTCACC	434	
336	ATCAGCAGACTGGAGCCCGGAAGATGTGGCACTTTATTACTGTACGCAATATTTTACTACT	395	
433	ATCACCAGAGTGGAGCCTGAAGACTTTGCACCTGTACTACTGTACAGGTCTATGGTGCCTCC	374	
396	CCGTACACTTTTGGCCAGGGACCAGGCTGGAGATCAAGGAACTGTGGCTGCACCATCT	455	
373	TCGTACACTTTTGGCCAGGGGACCAACTGGAGAGGAAACGAACTGTGCCTGCACCATCT	314	
456	GTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTCTGTGTGTGTC	515	
313	GTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTCTGTGTGTGTC	254	
516	CTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAAAGTGGATAACGCCCTC	575	
253	CTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAAAGTGGATAACGCCCTC	194	
576	CAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGCAGACACCTACAGC	635	
193	CAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGCAGACACCTACAGC	134	
636	CTCAGCAGCACCCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGC	695	
133	CTCAGCAGCACCCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGC	74	
696	GAAGTCACCCCATCAGGGCCCTGAGCTGCCCGTCAACAAAGAGCTTCAACAGGGGAGAGTGT	755	
73	GAAGTCACCCCATCAGGGCCCTGAGATGCCCGTCAACAAAGAGCTTCAACAGGGGAGAGTGT	14	
756	TAGAGGGAGA	765	
13	TAATTCTAGA	4	

## RESULT 13

PCT-US95-08743-156  
; Sequence 156, Application PC/TUS9508743  
: GENERAL INFORMATION

GENERAL INFORMATION:  
APPLICANT:

1	REFERENCE:	
2	TITLE OF INVENTION:	HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
3	TITLE OF INVENTION:	TO HUMAN IMMUNODEFICIENCY VIRUS
4	NUMBER OF SEQUENCES:	170
5	COMMENTS:	

COMPUTER READABLE FORM:

```

; : MEDIUM TYPE: Floppy disk
; :

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (FPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 807-4967

APPLICATION NUMBER: PCT/US95/08743  
FILING DATE: 11 - III - 1995

ESTIMATING DATE: II-JUL-1995  
PRIOR APPLICATION DATA.

APPLICATION NUMBER: US 08/275 953

FILING DATE: 18-JUL-1994

INFORMATION FOR SEQ ID NO: 156:

SEQUENCE CHARACTERISTICS:

LENGTH: 13254 base pair

TYPE: nucleic acid

STRADEDNESS: double

**TOPOLOGY:** circular

MOLECULE TYPE: DNA (genomic)

PCT-US95-08743-156

### Query Match

Query Match	58.8%;	Score 565.2;	DB 5;	Length 13254;
Best Local Similarity	86.8%;	Pred. No. 6.8e-152;		
Matches 634; Conservative	0;	Mismatches 93;	Indels 3;	Gaps 1;
QY 39	GACCCAGAGGAAACCATGGAAGCCCGCCAGCTCAGCTTCTCTCTCCCTCCTGCTACTCTGGCTC	98		
Db 12522	GACACGAAGCTTACCATTGGGTGTGCCCACTCAGGTCCTGGGTTGCTGCTGCTGGCTT	12581		
QY 99	CCAGATACCAACCGGAGAAATTGTGTTGACACAGTCTCCAGGCCACCTGTCTTTGTCCTCA	158		
Db 12582	ACAGATGCCAGATGTGAGATCGTTCTCACGCGAGTCTCCAGGCCACCTGTCTGTCTCTCA	12641		
QY 159	GGGGAAGAGCCACCCCTCTCTCCAGGGCCAGTCAGAGTGT---AGCAGCTACTTAGCC	215		
Db 12642	GGGGAAGAGCCACCTTCTCTGTAGGTCAGTCACAGCATTCGACGCCGCGCGGTAGCC	12701		
QY 216	TGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCAAACAGG	275		
Db 12702	TGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCAAAGCTGGTCATACATGGTGTTC	12761		
QY 276	GCCACTGGCATCCACACCCAGGTTCACTGGCAGTGGGTCTGGGACAGACTTCACTCTCACC	335		
Db 12762	GCCTCTGGCATCTCAGACAGGTTACGGGCGAGTGGGTCTGGGACAGACTTCACTCTCACC	12821		
QY 336	ATCAGCAGACTTGGCCAGGGGACCGAGTGTGGCAGCTTATTACTGTCTCAGCAATATTTACTACT	395		
Db 12822	ATCACCAGAGTGGAGCCTGAAGACTTTCAGCTGTACTACTGTTCAGGTTATGGTGCCTCC	12881		
QY 396	CCGTACACTTTTGGCCAGGGGACCGAGGCTGGAGATCAAAACGAACCTGTGGCTGCACCATCT	455		
Db 12882	TCGTACACTTTTGGCCAGGGGACCAAACTGGAGAGGAACGAACCTGTGCCTGCACCATCT	12941		
QY 456	GTCTTCATCTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAATCTGCCTCTGTTGTGTGC	515		
Db 12942	GTCTTCATCTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAATCTGCCTCTGTTGTGTGC	13001		
QY 516	CTGCTGAATAACTTCTATCCCGAGAGGCGCAAGTACAGTGGAAAGGTGGATAACGCCCTC	575		
Db 13002	CTGCTGAATAACTTCTATCCCGAGAGGCGCAAGTACAGTGGAAAGGTGGATAACGCCCTC	13061		
QY 576	CAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGC	635		
Db 13062	CAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGC	13121		
QY 636	CTCAGCAGCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGC	695		
Db 13122	CTCAGCAGCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGC	13181		
QY 696	GAAGTCAACCATCAGGGCCTGAGCTCGCCCGTCAAAAGAGCTTCAACAGGGGAGAGTGT	755		
Db 13182	GAAGTCAACCATCAGGGCCTGAGATCGCCCGTCAAAAGAGCTTCAACAGGGGAGAGTGT	13241		
QY 756	TAGAGGGAGA 765			
Db 13242	TAATTCTAGA 13251			

RESULT 14

PCT-US95-08743-170/C

; Sequence 170, Application PC/TUS9509743

GENERAL INFORMATION:

APPLICANT:

[illegible]

TITLE OF INVENTION

; NUMBER OF SEQUENCE  
; COMPILED BY:

COMPUTER READABLE  
MEDIUM

MEDIUM TYPE: FL  
COMPUTER: TR

COMPUTER: IBM PC  
OPERATING SYSTEM:

OPERATING SYSTEM

; SOFTWARE: Patent  
; CURRENT APPLICATION:



Tue Jul 29 09:34:11 2003

Db 241 GCATCTCAGACAGGTTTCAGCGGAGTGGTCTGGACAGACTTCACTCTCACCATCACCA 300  
QY 343 GACTGGAGCCCGAAGATGTGGCACTTTATTACTGTACAGCAATATTTTACTACTCCGTACA 402  
Db 301 GAGTGGAGCCTGAAGACTTTTGCACTGTACTACTGTCAAGGTCTATGGTGCCTCCTCGTACA 360  
QY 403 CTTTGGCCAGGGACAGGCTGGAGATCAAAACGAACTGTGGCTGCACCATCTGTCTTCA 462  
Db 361 CTTTGGCCAGGGACCAAACTGGAGAGAAACGAACTGTGGCTGCACCATCTGTCTTCA 420  
QY 463 TCTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAACCTGCTCTGTTGTGCTGCTGA 522  
Db 421 TCTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAACCTGCTCTGTTGTGCTGCTGA 480  
QY 523 ATAACTTCTATCCCGAGAGGCGCCAAAGTACAGTGAAGGTGGATAACGCCCTCCAATCGG 582  
Db 481 ATAACTTCTATCCCGAGAGGCGCCAAAGTACAGTGAAGGTGGATAACGCCCTCCAATCGG 540  
QY 583 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGCCTCAGCA 642  
Db 541 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGCCTCAGCA 600  
QY 643 GCACCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCCAAGTCA 702  
Db 601 GCACCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCCAAGTCA 660  
QY 703 CCCATCAGGGCCTGAGCTCGCCCGTCAAAAGAGCTTCAACAGGGGAGGTGTAGAGGG 762  
Db 661 CCCATCAGGGCCTGAGTTCGCCCGTCAAAAGAGCTTCAACAGGGGAGGTGTATTTCT 720  
QY 763 AGA 765  
Db 721 AGA 723

Search completed: July 25, 2003, 19:56:46  
Job time : 51.6192 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 14:13:55 ; Search time 2529.23 Seconds  
(without alignments)  
11069.353 Million cell updates/sec

Title: US-09-806-276A-3  
Perfect score: 962  
Sequence: 1 tcgagccgattcggtcgag.....gaatctttgcacaaaaaaaaa 962

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	870	90.4	1020	9	BC030813	BC030813 Homo sapi
2	857	89.1	968	9	BC016380	BC016380 Homo sapi
3	845.4	87.9	970	6	AX305008	AX305008 Sequence
4	845.4	87.9	970	6	AX306537	AX306537 Sequence
5	836.8	87.0	997	9	BC032451	BC032451 Homo sapi
6	830.2	86.3	948	6	AX305004	AX305004 Sequence
7	830.2	86.3	948	6	AX306533	AX306533 Sequence
8	829.6	86.2	931	6	ARI35361	ARI35361 Sequence
9	823.2	85.6	924	6	E07333	E07333 cdNA sequen
10	792.4	82.4	1244	6	AX333289	AX333289 Sequence
11	792.4	82.4	1244	9	HUMIGGK	M63438 Human Ig re
12	756.4	78.6	956	9	BC029444	BC029444 Homo sapi
13	753.2	78.3	962	9	BC034141	BC034141 Homo sapi
14	751	78.1	3819	6	AR161402	AR161402 Sequence
15	746.8	77.6	953	9	BC005332	BC005332 Homo sapi
16	746.8	77.6	974	6	AX305000	AX305000 Sequence
17	746.8	77.6	974	6	AX306529	AX306529 Sequence
18	740.4	77.0	938	9	HSA010442	AJ010442 Homo sapi
19	735.6	76.5	944	6	AX067344	AX067344 Sequence
20	734	76.3	938	9	HSIGGK1L	Y14736 Homo sapien
21	734	76.3	961	9	BC034146	BC034146 Homo sapi
22	726	75.5	944	9	BC018761	BC018761 Homo sapi
23	716.4	74.5	979	9	BC017870	BC017870 Homo sapi
24	710.4	73.8	928	6	AX251598	AX251598 Sequence
25	704.8	73.3	929	9	AK097353	AK097353 Homo sapi
26	701.2	72.9	977	9	BC034142	BC034142 Homo sapi
27	695	72.2	972	9	BC030814	BC030814 Homo sapi
28	691.8	71.9	973	9	BC022362	BC022362 Homo sapi
29	685	71.2	958	6	AR135357	AR135357 Sequence
30	683.6	71.1	802	6	AX067324	AX067324 Sequence
31	683.2	71.0	1952	9	AK024974	AK024974 Homo sapi
32	680	70.7	945	6	A21386	A21386 Plasmid DNA
33	680	70.7	963	6	AX067347	AX067347 Sequence
34	678.6	70.5	944	9	AK096938	AK096938 Homo sapi
35	656.8	68.3	708	6	AX419497	AX419497 Sequence
36	615.6	64.0	708	9	AF027158	AF027158 Homo sapi
37	601	62.5	811	9	AB064068	AB064068 Homo sapi
38	600	62.4	891	6	E40748	E40748 Antihuman F
39	599.4	62.3	830	9	AB064109	AB064109 Homo sapi
40	599.2	62.3	794	9	AB064129	AB064129 Homo sapi
41	599	62.3	908	9	AF186176	AF186176 Homo sapi
42	596	62.0	781	9	AB064131	AB064131 Homo sapi
43	592.8	61.6	807	9	AB064071	AB064071 Homo sapi
44	592.6	61.6	833	9	AB064110	AB064110 Homo sapi
45	591.4	61.5	738	9	AB064124	AB064124 Homo sapi

ALIGNMENTS

RESULT 1  
BC030813  
LOCUS BC030813 1020 bp mRNA linear PRI 13-JUN-2002  
DEFINITION Homo sapiens, clone MGC:22645 IMAGE:4700961, mRNA, complete cds.  
ACCESSION BC030813  
VERSION BC030813.1 GI:21410095  
KEYWORDS MGC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1020)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 37 Row: a Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

source

Location/Qualifiers  
1..1020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:22645 IMAGE:4700961"  
/tissue\_type="Lung"  
/clone\_lib="NIH\_MGC\_77"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
34..738  
/codon\_start=1  
/product="Unknown (protein for MGC:22645)"  
/protein\_id="AAH30813.1"  
/db\_xref="GI:21410096"  
/translation="MEAPAQLLFLLLWLPDITGEIVMTQSPATLSVSPGERATLSCR  
ASQVTSNLAWYQQTGSPRLVIYGASSRASGVPARFSGSGTEFTLISLSQSED  
FAVYVQYQYKWPHTFGQTKLDIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFY  
PREAKVQWKVDNALQSGNSQESVTEQDSKDSYLSSTLTLSKADYKEKHVYACEVTH  
QGLSPVTKSFNRGEC"

CDS

BASE COUNT 309 a 282 c 227 g 202 t  
ORIGIN

Query Match 90.4%; Score 870; DB 9; Length 1020;  
Best Local Similarity 95.7%; Pred. No. 3.4e-221;  
Matches 894; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY	29	GCTCAGTTAGGACCCAGAGGGAACCATGGAGCCCGCCAGCTCAGCTTCTCTTCTCTCTGCT	88
Db	9	GCTCAGTTAGGACCCAGAGGGAACCATGGAGCCCGCCAGCTTCTCTTCTCTCTGCT	68
QY	89	ACTCTGGCTCCCGATACCCACCGGAGAAATTGTTGACACAGTCTCCAGCCACCTGTC	148
Db	69	ACTCTGGCTCCCGATACCCACCGGAGAAATTGTTGACACAGTCTCCAGCCACCTGTC	128
QY	149	TTTGTCTCCAGGGGAAAGAGCCACCTCTCTCTGAGGCGCAGTCCAGAGTGTAGCAGCTA	208
Db	129	TGTGTCTCCAGGGGAAAGAGCCACCTCTCTCTGAGGCGCAGTCCAGAGTGTAGCAGCA	188
QY	209	CTTAGCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGTCATC	268
Db	189	CTTAGCTGGTACCAACAGAAACCTGGCCAGGCTCCTCATCTATGTCATCTATGTCATC	248
QY	269	CAACAGGCGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCAC	328
Db	249	CAGCAGGCGCCAGTGGTGTCCAGCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCAC	308
QY	329	TCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTATTACTGTACGAATATTT	388
Db	309	TCTCACCATCAGCAGCTGCGAGTCTGAAGATTTTGCAGTTTATTACTGTACGAATATAA	368

QY	389	TACTACTCCGTACACTTTTGGCCAGGGGACCAGGCTGGAGATCAAAACGAACCTGTGCTGC	448
Db	369	TAAAGTGGCCGACACACTTTTGGCCAGGGGACCAGGCTGGAGATCAAAACGAACCTGTGCTGC	428
QY	449	ACCATCTGTCTTCTATCTTCCCGCCATCTGATGAGCAGTGAATCTGGAACCTGCTCTGT	508
Db	429	ACCATCTGTCTTCTATCTTCCCGCCATCTGATGAGCAGTGAATCTGGAACCTGCTCTGT	488
QY	509	TGTGTGCTGCTGCTGAATAACTTCTATCCAGAGAGGGCCAAAGTACAGTGGAGGTGGATAA	568
Db	489	TGTGTGCTGCTGCTGAATAACTTCTATCCAGAGAGGGCCAAAGTACAGTGGAGGTGGATAA	548
QY	569	CGCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC	628
Db	549	CGCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC	608
QY	629	CTACAGCCTCAGCAGCACCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTA	688
Db	609	CTACAGCCTCAGCAGCACCTGACGCTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTA	668
QY	689	CGCCTGCGAAGTCAACCCATCAGGGCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGG	748
Db	669	CGCCTGCGAAGTCAACCCATCAGGGCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGG	728
QY	749	AGAGTGTAGAGGGAGAGTACCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCA	808
Db	729	AGAGTGTAGAGGGAGAGTACCTCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCA	788
QY	809	TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCCTCCAGCTC	868
Db	789	TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCCTCCAGCTC	848
QY	869	ATCTTTACCTCACCCCT	928
Db	849	ATCTTTACCTCACCCCT	908
QY	929	TGAATAAATAAAGTGAATCTTTTGCAAAAAA 962	
Db	909	TGAATAAATAAAGTGAATCTTTTGCAAAAAA 942	

RESULT 2  
BC016380  
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC016380  
Homo sapiens, similar to immunoglobulin kappa constant, clone  
MGC:27376 IMAGE:4688477, mRNA, complete cds.  
BC016380  
MGC.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 968)  
Straussberg, R.  
Direct Submission  
Submitted (29-OCT-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
Clone distribution: MGC clone distribution information can be found

QY CACCTACAGCCTTCAGCAGCCACCTTGACGCTGAGCAAGCAGATCACCGTGTGGTTTCTG

by  
104 GRANTING CONCERN











```
sig_peptide      28. .87
3'UTR            739. .948
BASE COUNT      231 a   285 c   224 g   207 t   1 others
ORIGIN
Query Match      86.3%;   Score 830.2;   DB 6;   Length 948;
Best Local Similarity 96.1%;   Pred. No. 1.4e-210;
Matches 873;   Conservative 0;   Mismatches 29;   Indels 6;   Gaps 2;

QY  52 CCATGGAAGCCCCAGCTCAGCTTCTCTCTCCTCCTGCTACTCTGGCTCCAGATACCACCG 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  26 CCATGGAACCCAGCGCAGCTTCTCTCCTCCTGCTACTCTGGCTCCAGATACCACCG 85

QY  112 GAGAAATGTGTTGACACAGTCTCCAGCCACCCCTGCTTTGTCTCCAGGGGAAAGAGCCA 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  86 GAGAAATGTGTTGACGCACTCTCCAGGCACCCCTGCTTTGTCTCCAGGGGAAAGAGCCA 145

QY  172 CCCTCTCCTGCAGGGCCAGTCAGAGTGT---AGCAGCTACTTAGCCTGGTACCAACAGA 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  146 CCCTCTCCTGCAGGGCCAGTCAGAAATATTAGAAGCAGCTACTTAGCCTGGTACCAGCAGA 205

QY  229 AACCTGGCCAGGCTCCCAGGCTCCTCATCTATGATGATCCCAACAGGGCCACTGGCATCC 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  206 AACCTGGCCAGGCTCCCGGCTCCTCATCTATGTTGCATCCAGCAGGGCCACTGGCATCC 265

QY  289 CACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTCAOATCAGCAGACTGG 348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  266 CAGACAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTCAOATCAGCAGACTGG 325

QY  349 AGCCGGAAGATGTGGCACTTTATTACTGTCTCAGCAATATTTTA---CTACTCGGTACACTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  326 AGCCTGAAGATTTTGCAGTGTATTACTGTTCAGCAGTTTGGTAGCTCACCTATGTGCAGTT 385

QY  406 TTGGCAGGGGACCAGGCTGGAGATCAAAACGAACTGCGGTGCACCATCTGCTTCATCT 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  386 TTGGCAGGGGACCAGCTGGAGATCAAAACGAACTGCGGTGCACCATCTGCTTCATCT 445

QY  466 TCCCGCATCTGATGAGCAGTGAATCTGGAACCTGCTCTGTTGTGCTGCTGAATA 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  446 TCCCGCATCTGATGAGCAGTGAATCTGGAACCTGCTCTGTTGTGCTGCTGAATA 505

QY  526 ACTTCTATCCAGAGAGGCGCAAGTACAGTGGAGGTGGATAACGCCCTCCCAATCGGGTA 585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  506 ACTTCTATCCAGAGAGGCGCAAGTACAGTGGAGGTGGATAACGCCCTCCCAATCGGGTA 565

QY  586 ACTCCAGGAGAGTGTACACAGCAGGACAGCAAGCAGCACCTACAGCCTCAGCAGCA 645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  566 ACTCCAGGAGAGTGTACACAGCAGGACAGCAAGCAGCACCTACAGCCTCAGCAGCA 625

QY  646 CCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAAGTCTACGCCCTGCGAAGTCACCC 705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  626 CCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAAGTCTACGCCCTGCGAAGTCACCC 685

QY  706 ATCAGGGCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGAGA 765
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  686 ATCAGGGCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGAGA 745

QY  766 AGTGCCCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCCTTTGGCCTCTGACC 825
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  746 ANTGCCCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCCTTTGGCCTCTGACC 805

QY  826 CTTTTCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTCATCTTACCTCACCCC 885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  806 CTTTTCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTCATCTTACCTCACCCC 865

QY  886 CCTCTCCTCTTGGCTTTAATTATGCTAATGTTGAGGAGAGTGAATAAATAAAGTGAA 945
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  866 CCTCTCCTCTTGGCTTTAATTATGCTAATGTTGAGGAGAGTGAATAAATAAAGTGAA 925

QY  946 TCCTTGCA 953
    ||||| |||||
Db  926 TCCTTGCA 933
    ||||| |||||
```

```
RESULT 8
AR135361
LOCUS          AR135361          931 bp      DNA      linear      PAT 16-JUN-2001
DEFINITION     Sequence 19 from patent US 6135941.
ACCESSION      AR135361
VERSION        AR135361.1  GI:14476033
KEYWORDS
SOURCE         Unknown.
ORGANISM       Unknown.
REFERENCE      1 (bases 1 to 931)
AUTHORS       Hillman,J.L., Lal,P., Tang,Y.Tom., Yue,H., Au-Young,J.,
               Corley,N.C., Guegler,K.J. and Baughn,M.R.
TITLE         Human immune system associated molecules
JOURNAL       Patent: US 6135941-A 19 24-OCT-2000;
FEATURES       Location/Qualifiers
               source
               1..931
               /organism="unknown"
BASE COUNT     224 a   283 c   218 g   206 t
ORIGIN
Query Match    86.2%;   Score 829.6;   DB 6;   Length 931;
Best Local Similarity 93.6%;   Pred. No. 2e-210;
Matches 865;   Conservative 0;   Mismatches 59;   Indels 0;   Gaps 0;

QY  39 GACCCAGAGGGAACCATGGAAGCCCCCAGCTCAGCTTCTCTCCTCCTGCTACTCTGGCTC 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1  GACCCAGAGGGAACCATGGAAGCCCCCAGCGCAGCTTCTCTCCTCCTGCTACTCTGGCTC 60

QY  99 CCAGATACCCACGGGAGAAATTGTGTTGACACAGTCTCCAGCCACCCTGTCTTTGTCTCCA 158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  61 CCAGATTCTTTGGAGAAATTGTAATGAGCGAGTCTCCAGCCAGTCTGTCTCTCTCCA 120

QY  159 GGGGAAAGAGCCACCCCTCTCCTGCGAGGCCAGTCAGAGTGTAGCAGCTACTTAGCCTGG 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  121 GGGGAAAGAGCCACCCCTTCTCCTGCGAGGCCAGTCAGTCTATTGGCAGCACCATCGCCTGG 180

QY  219 TACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCAACAGGGCC 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  181 TACCAACAAAGACCTGGCCAGTCTCCAGGCTCCTCATCTATGCTTCCACCCGGGCC 240

QY  279 ACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACCTCTCACCATC 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  241 ACTGGTGTCCACCCAGGTTTCAGTGGCGGTGGGTCTGGGACAGAGTTCACCTCTCTTCATC 300

QY  339 AGCAGACTGGAGCCCGGAAGATGTGGCACTTTATTACTGTGCAAGATATTTTACTACTCCG 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  301 AGCAGCCTGCAGTCTGAAGATTTTGCACCTTTATTACTGTGCAAGATATAAAGGTGGCCG 360

QY  399 TACACTTTTGGCCAGGGGACCAGGCTGGAGATCAAAACGAACCTGTGGCTGCACCATCTGTC 458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  361 CTCACCTTTCGGCGGAGGGACCAGGTTGCAGATCAAAACGAACCTGTGGCTGCACCATCTGTC 420

QY  459 TTCATCTTCCGCCCATCTGATGAGCAGTTGAAATCTGGAATCTGGAATCGCTCTGTTGTGCCCTG 518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  421 TTCATCTTCCGCCCATCTGATGAGCAGTTGAAATCTGGAATCTGGAATCGCTCTGTTGTGCCCTG 480

QY  519 CTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATTAACGCCCTCCAA 578
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  481 CTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATTAACGCCCTCCAA 540

QY  579 TCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGACACACCTACAGCCTC 638
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  541 TCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGACACACCTACAGCCTC 600

QY  639 AGCAGCACCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAA 698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  601 AGCAGCACCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAA 660

QY  699 GTCACCCATCAGGGGCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAG 758
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  661 GTCACCCATCAGGGGCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAG 720
```



QY 759 AGGAGAGAGTCCCCACCTGCTCCTCAGTTCAGCTCCAGCCTGACCCCTCCCATCCTTTGGCC 818  
|||||  
Db 721 AGGAGAGAGTCCCCACCTGCTCCTCAGTTCAGCTCCAGCCTGACCCCTCCCATCCTTTGGCC 780  
|||||  
QY 819 TCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCCACC 878  
|||||  
Db 781 TCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCCACC 840  
|||||  
QY 879 TCACCCCTCCTCCTCCTCCTTTGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAAATA 938  
|||||  
Db 841 TCACCCCTCCTCCTCCTCCTTTGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAAATA 900  
|||||  
QY 939 AAGTGAATCTTTGCAAAAAA 962  
|||||  
Db 901 AAGTGAATCTTTGCAAAAAA 924  
|||||

RESULT 9  
E07333 LOCUS  
DEFINITION  
E07333 CDNA sequence coding the amino acid sequence of L-chain of the antibody inhibiting histamine elimination.  
ACCESSION  
E07333.1 GI:2175473  
VERSION  
JP 1994113881-A/1.  
KEYWORDS  
unidentified.  
SOURCE  
unidentified.  
ORGANISM  
unclassified.  
REFERENCE  
1 (bases 1 to 924)  
AUTHORS  
Washida,N., Yoshida,N., Morinaga,T., Mizuno,A., Goto,M. and Kobayashi,F.  
TITLE  
HUMAN TYPE MONOCLONAL ANTIPEPTIDE ANTIBODY AND DNA CAPABLE OF CODING THE SAME  
JOURNAL  
Patent: JP 1994113881-A 1 26-APR-1994;  
COMMENT  
SNOW BRAND MILK PROD CO LTD  
OS Unclassified  
PN JP 1994113881-A/1  
PD 26-APR-1994  
PF 07-OCT-1992 JP 1992293800  
PI WASHIDA NAOHIRO, YOSHIDA NORIKO, MORINAGA TOMONORI, PI MIZUNO ATSUKO,  
PI GOTO MASAOKI, KOBAYASHI FUMIE  
PC C12P21/08,A61K39/395,C07K13/00,C12N15/07,C12N15/13,G01N33/53,  
PC G01N33/577//  
PC C12N5/24,(C12P21/08,C12R1:91),(C12N5/24,C12R1:91); CC  
strandedness: Single;  
CC topology: Linear;  
FH Key  
FH Location/Qualifiers  
FT source 1..924  
FT /organism='Unclassified'.  
FT Location/Qualifiers  
FT 1..924  
FT /organism='unidentified'  
FT /db\_xref='taxon:32644'

BASE COUNT 230 a 282 c 214 g 195 t 3 others  
ORIGIN

Query Match 85.6%; Score 823.2; DB 6; Length 924;  
Best Local Similarity 93.4%; Pred. No. 1e-208;  
Matches 858; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 44 AGAGGAACCATGGAAGCCAGCTCAGCTTCTCTCTCCTCCTGCTACTCTGGCTCCCGA 103  
|||||  
Db 1 AGACGAACCATGGAAGCCAGCGCAGCTTCTCTCTCCTCCTGCTACTCTGGCTCCCGA 60  
|||||  
QY 104 TACACGGGAGAAATGTGTGACACAGTCTCCAGCCACCTGTCTTTGTCTCCAGGGGA 163  
|||||  
Db 61 TACCACTGGAGAAATAGTGATGACGCGAGTCTCCAGCCACCTGTCTGTGTCTCCAGGGG 120  
|||||  
QY 164 AAGAGCACCTCTCCTGCGAGGCCAGTCCAGTGTAGCAGCTACTTAGCCTGGTACCA 223  
|||||

Db 121 AAGAGCCGCCCTCTCCTGCAAGCCAGTCCAGAGTGTAGCAACAACATAGCCTGGTACCA 180  
|||||  
QY 224 ACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGATCAACAGGGCCACTGG 283  
|||||  
Db 181 GCAGAAACCTGCCAGGCTCCAGGCTCCTCATCTATGGTGGCTCCAGGGCCACTGG 240  
|||||  
QY 284 CATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAG 343  
|||||  
Db 241 TATCCCGCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCACCAC 300  
|||||  
QY 344 ACTGGAGCCCGAAGATGTGGCACTTTATTACTGTGAGCAATATTTACTACTCCGTACAC 403  
|||||  
Db 301 CCTACAGTCTCAACATTTTGGCAATTTATTACTGTGAGCAATATATAGTAGTGGCCTCGGAC 360  
|||||  
QY 404 TTTTGGCCAGGGACCGAGGCTGGAGATCAAAACGAACTGTGGTGCACCATCTGTCTTCAT 463  
|||||  
Db 361 GTTCGGCCAGGGACCAAGGTGGACCTCAAGGAACTGTGGTGCACCATCTGTCTTCAT 420  
|||||  
QY 464 CTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCTGCCCTCTGTGTGCTGCTGAA 523  
|||||  
Db 421 CTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCTGCCCTCTGTGTGCTGCTGAA 480  
|||||  
QY 524 TAACCTTCTATCCAGAGAGAGGCGCAAGTACAGTGGAGGTGATGAACGCCCTCCAAATCGGG 583  
|||||  
Db 481 TAAGTCTATCCAGAGAGAGGCGCAAGTACAGTGGAGGTGGATAACGCCCTCCAAATCGGG 540  
|||||  
QY 584 TAACCTCCAGAGAGAGTGTACAGAGCAGGACAGCAAGGACAGCCTACAGCCTCAGCAG 643  
|||||  
Db 541 TAACCTCCAGAGAGAGTGTACAGAGCAGGACAGCAAGGACAGCCTACAGCCTCAGCAG 600  
|||||  
QY 644 CACCCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCAC 703  
|||||  
Db 601 CACCCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTGTACGCCCTGCGAAGTCAC 660  
|||||  
QY 704 CCATCAGGGCCTGAGCTCGCCCTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGA 763  
|||||  
Db 661 CCATCAGGGCCTGAGCTCGCCCTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGA 720  
|||||  
QY 764 GAAGTGGCCCCACCTGCTCCTCAGTTCAGCTCCAGCCTGACCCCTCCCATCCTTTGGCCTCTGA 823  
|||||  
Db 721 GAAGTGGCCCCACCTGCTCCTCAGTTCAGCTCCAGCCTGACCCCTCCCATCCTTTGGCCTCTGA 780  
|||||  
QY 824 CCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCCACCTACC 883  
|||||  
Db 781 CCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCCACCTACC 840  
|||||  
QY 884 CCCTCCTCCTCCTTGGCTTTAATATGCTAATGTTGGAGGAGAAATGAATAAAGTG 943  
|||||  
Db 841 CCCTCCTCCTCCTTGGCTTTAATATGCTAATGTTGGAGGAGAAATGAATAAAGTG 900  
|||||  
QY 944 AATCTTTGCAAAAAA 962  
|||||  
Db 901 AATCTTTNNAAAAA 919  
|||||

RESULT 10  
AX333289

LOCUS  
DEFINITION  
AX333289 Sequence 3798 from Patent WO0194629.

ACCESSION  
AX333289

VERSION  
AX333289.1 GI:18123923

KEYWORDS  
human.

SOURCE  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,

TITLE  
Horrigan,S., Soppet,D.R. and Weaver,Z.  
Cancer gene determination and therapeutic screening using signature

JOURNAL  
gene sets  
Patent: WO 0194629-A 3798 13-DEC-2001;  
Avalon Pharmaceuticals (US)

AX333289 1244 bp DNA linear PAT 09-JAN-2002



FEATURES		Location/Qualifiers	
source		1..1244	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
BASE COUNT	292 a	383 c	302 g 267 t
ORIGIN			
Query Match			
Best Local Similarity 82.4%; Score 792.4; DB 6; Length 1244;			
Matches 827; Conservative 0; Mismatches 31; Indels 6; Gaps 1;			
QY	95	GCTCCAGATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGCTTTGTC	154
Db	381	GATCAACATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGCTTTGTC	440
QY	155	TCCAGGGGAAAGAGCCACCTCTCTGCAGGGCCAGTCAGAGTGTAGCAGCTACTTAGC	214
Db	441	TCCAGGGGAAAGAGCCACCTCTCTGCAGGGCCAGTCAGAGTGTAGCAGCTACTTAGC	500
QY	215	CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCATCTATGATGATCAACACAG	274
Db	501	CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCATCTATGATGATCAACACAG	560
QY	275	GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTAC	334
Db	561	GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTAC	620
QY	335	CATCAGCAGACTGGAGCCCGAAGATGTGGCAGTCTTATTACTGTGAGC-----AAATATT	388
Db	621	CATCAGCAGCTAGAGCCTGAAGATTGTGCAGTTTATTACTGTCAACACCCGTGACAATTG	680
QY	389	TACTACTCCGTACACTTTTGGCCAGGGACAGGCTGGAGATCAACAGAACTGTGGCTGC	448
Db	681	GCCTCCGGGGCCACTTTTCGGCGGAGGACCAAGTGGAGATCAACAGAACTGTGGCTGC	740
QY	449	ACCATCTGTCTTCACTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCCTCTGT	508
Db	741	ACCATCTGTCTTCACTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCCTCTGT	800
QY	509	TGTGTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGTGGATAA	568
Db	801	TGTGTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGTGGATAA	860
QY	569	CGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAGCAC	628
Db	861	CGCCCTCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAGCAC	920
QY	629	CTACAGCCTCAGCAGCACCCCTGACGTGAGCAAGCAGACTACGAGAAACACAAAGTCTA	688
Db	921	CTACAGCCTCAGCAGCACCCCTGACGTGAGCAAGCAGACTACGAGAAACACAAAGTCTA	980
QY	689	CGCCTGCGAAGTCAACCCATCAGGCGCTGAGCTCGCCCGTCAAAAGAGCTTCAACAGGG	748
Db	981	CGCCTGCGAAGTCAACCCATCAGGCGCTGAGCTCGCCCGTCAAAAGAGCTTCAACAGGG	1040
QY	749	AGAGTGTAGAGGGAGAAAGTGGCCCGCCACCTGCTCCTCAGTTCAGCCCTGCCCA	808
Db	1041	AGAGTGTAGAGGGAGAAAGTGGCCCGCCACCTGCTCCTCAGTTCAGCCCTGCCCA	1100
QY	809	TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTC	868
Db	1101	TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTC	1160
QY	869	ATCTTTACCTCACCCCTCCCTCTCTTGGCTTTAATTAAGTGTGGAGGAGAA	928
Db	1161	ATCTTTACCTCACCCCTCCCTCTCTTGGCTTTAATTAAGTGTGGAGGAGAA	1220
QY	929	TGAATAAATAAGTGAATCTTTGC	952
Db	1221	TGAATAAATAAGTGAATCTTTGC	1244
RESULT 11			

HUMIGGK																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

HUMIGGK	Human Ig rearranged gamma chain mRNA, V-J-C region and complete cds.		1244 bp	mRNA	linear	PRI 27-APR-1993
LOCUS						
DEFINITION						
ACCESSION	M63438					
VERSION	M63438.1	GI:184847				
KEYWORDS	C-region; V-region; immunoglobulin gamma-chain; immunoglobulin light chain.					
SOURCE	Human, cDNA to mRNA (transfected into mouse myeloma cell line P3X63Ag8.653).					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Shuford, W., Raff, H.V., Finley, J.W., Esselstyn, J. and Harris, L.J.					
TITLE	Effect of light chain V region duplication on IgG oligomerization and in vivo efficacy					
JOURNAL	Science 252 (5006), 724-727 (1991)					
MEDLINE	91220077					
PUBMED	1902593					
FEATURES	Location/Qualifiers					
source	1..1244					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
CDS	join(1..49,50..388,389..727,728..1050)					
	/codon_start=1					
	/protein_id="AAA73002.1"					
	/db_xref="GI:184848"					
	/translation="MEAPAQLLFLLLWLPDPTTGEIVLTQSPATLSLSPGERATLSCR					
	ASQVGSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGGTDTLTISSELEP					
	FAVYVCOHRDNWPPGATFGGTKEIKHTTGEIVLTQSPATLSLSPGERATLSCRASQ					
	SVGSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGGTDTLTISSELEPFAV					
	YVCOHRDNWPPGATFGGTKEIKRTVAAPSVFIFPPSDEQLKSGTASVVCCLNNFYP					
	REAKVQWKVDNALQSGNSQESVTEQDSKDSYSLSTLTLSKADYEKHKVYACEVTHQ					
	GLSSPVTKSFNRGEC"					
sig_peptide	1..60					
exon	<1..49					
	/note="encodes first part of signal peptide"					
exon	50..388					
	/note="encodes second part of signal peptide and VJ region"					
mat_peptide	/number=2					
exon	61..1047					
	389..727					
	/note="duplication of exon 2"					
	/number=3					
exon	728..1244					
	/note="encodes human Kappa constant region"					
	/number=4					
polyA_signal	1223..1232					
BASE COUNT	292 a 383 c 302 g 267 t					
ORIGIN	chromosome 2.					
Query Match	82.4%; Score 792.4; DB 9; Length 1244;					
Best Local Similarity	95.7%; Pred. No. 1.7e-200;					
Matches 827; Conservative	0; Mismatches 31; Indels 6; Gaps 1;					
QY	95	GCTCCAGATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGCTTTGTC	154			
Db	381	GATCAACATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGCTTTGTC	440			
QY	155	TCCAGGGGAAAGAGCCACCTCTCTCCAGGGCCAGTCAGAGTGTAGCAGCTACTTAGC	214			
Db	441	TCCAGGGGAAAGAGCCACCTCTCTCCAGGGCCAGTCAGAGTGTAGCAGCTACTTAGC	500			
QY	215	CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCATCTATGATGATCAACACAG	274			
Db	501	CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCATCTATGATGATCAACACAG	560			
QY	275	GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTAC	334			
Db	561	GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTAC	620			



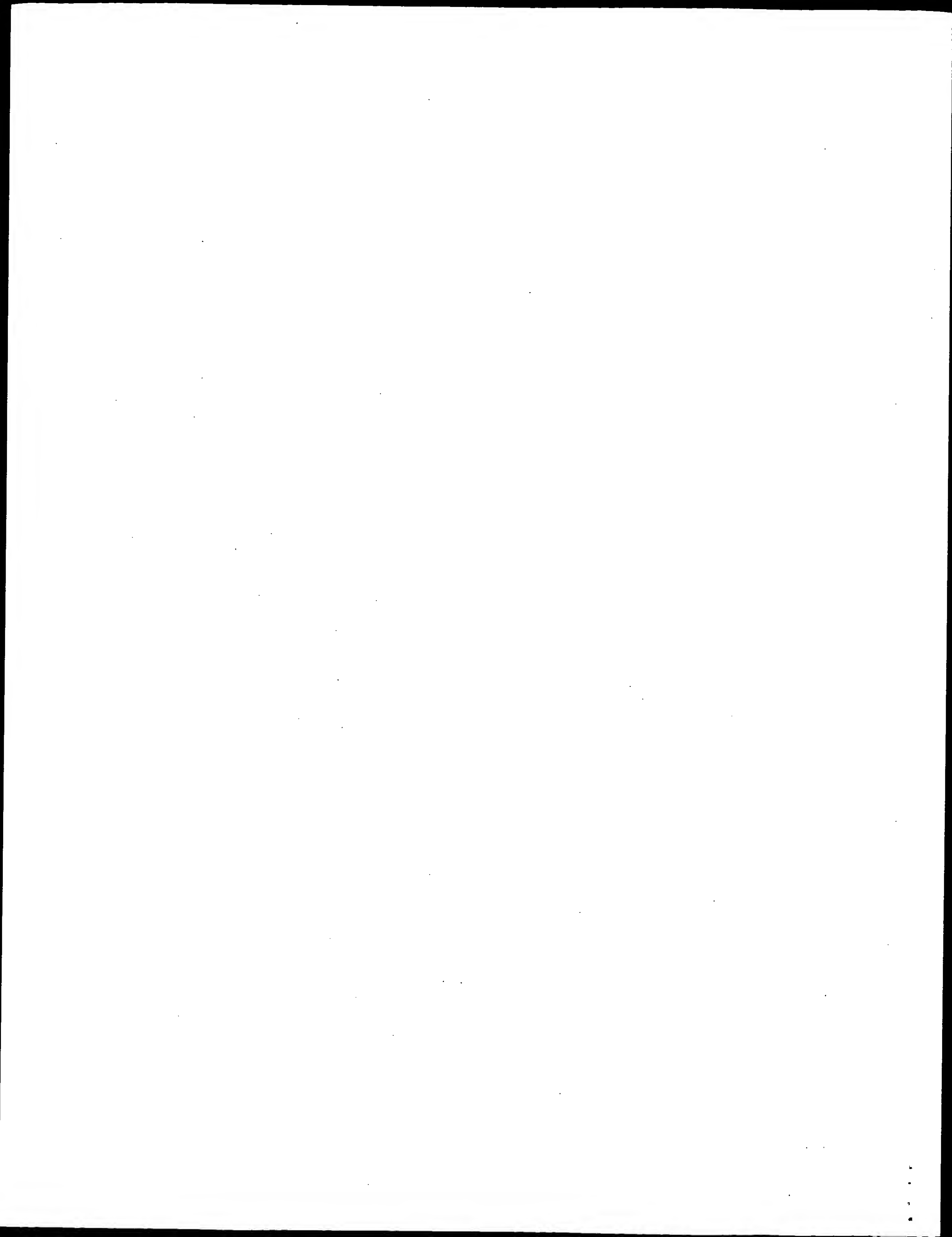












GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 25, 2003, 13:53:40 ; Search time 221.852 Seconds  
(without alignments)  
9765.155 Million cell updates/sec

Title: US-09-806-276A-3  
Perfect score: 962  
Sequence: 1 tcgagcgcattcgctcgag.....gaatctttgcaaaaaaaaaa 962

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	962	21	AAA09154 Human BMDSP-1 codi
2	880.4	91.5	1061	22	AAH31379 Human secreted pro
3	845.4	87.9	970	24	AAS99477 Anti-human AILIM m
4	839.6	87.3	924	15	AAQ71872 Sequence coding hu
5	830.2	86.3	948	24	AAS99475 Anti-human AILIM m
6	829.6	86.2	931	22	AAC66524 Human immune syste
7	808.4	84.0	913	24	AAS62744 CDNA sequence #531
8	792.8	82.4	1450	22	AAS22766 Human cDNA encodin
9	792.8	82.4	1450	22	AAS22767 Human cDNA encodin

10	792.8	82.4	1450	22	AAS22768 Human cDNA encodin
11	792.8	82.4	1450	22	AAS22769 Human cDNA encodin
12	792.4	82.4	1242	12	AAQ11878 lB1 IgG aberrant l
13	792.4	82.4	1244	24	ABK64543 Human benign prost
14	792.4	82.4	1244	24	ABL65461 Lung cancer relate
15	771.2	80.2	1721	22	AAS22530 Human cDNA encodin
16	771.2	80.2	1736	22	AAS22532 Human cDNA encodin
17	771.2	80.2	1741	22	AAS22531 Human cDNA encodin
18	761.6	79.2	1106	24	ABQ54241 Human ovarian anti
19	751	78.1	3819	18	AAT78825 Kappa light chain
20	751	78.1	3819	19	AAV39266 Plasmid pLC6G5 nuc
21	751	78.1	3819	20	AAZ22020 Nucleotide sequenc
22	748	77.8	936	21	AAA27390 Human IGFAM-10 imm
23	746.8	77.6	974	24	AAS99473 Anti-human AILIM m
24	738.8	76.8	1033	22	AAS00145 Human cDNA clone H
25	738.8	76.8	1066	14	AAQ49943 Human anti-HBs lig
26	735.6	76.5	944	22	AAF44892 Human breast cance
27	727	75.6	917	21	AAA27381 Human IGFAM-1 immu
28	725.8	75.4	1517	22	AAH98191 Human EST-derived
29	724.6	75.3	990	22	AAF97952 Human secreted pro
30	723.6	75.2	930	22	AAH31350 Human secreted pro
31	721.4	75.0	928	21	AAA27393 Human IGFAM-13 imm
32	718.4	74.7	1404	22	AAS22861 Human cDNA encodin
33	717.6	74.6	923	21	AAA95785 Human immune syste
34	713.2	74.1	941	22	AAF97954 Human immune respo
35	710.4	73.8	928	22	AAH26799 Human IGFAM-5 immu
36	709.8	73.8	916	21	AAA27385 Human EST-derived
37	705	73.3	851	22	AAH98251 Human immune syste
38	685	71.2	958	22	AAC66520 Human breast cance
39	683.6	71.1	802	22	AAF44872 Encodes light chai
40	680	70.7	945	13	AAQ20067 Human breast cance
41	680	70.7	963	22	AAF44895 chiT84.12 l6 light
42	662.2	68.8	957	15	AAQ54654 Human IGFAM-12 imm
43	658.2	68.4	890	21	AAA27392 CDNA of the light
44	656.8	68.3	708	24	AAK98702 DNA encoding the k
45	652.2	67.8	705	21	AAA46871

ALIGNMENTS

RESULT 1  
AAA09154  
ID AAA09154 standard; DNA; 962 BP.

XX  
AC AAA09154;

XX  
DT 10-AUG-2000 (first entry)

XX  
DE Human BMDSP-1 coding sequence.

XX  
KW Bone marrow-derived serum protein; immunoglobulin kappa light chain;

KW BMDSP-1; cytostatic; anti-HIV; antisthmatic; antiatherosclerotic;

KW anti-inflammatory; antiarthritic; antisclerotic; hypotensive; antiviral;

KW antiparasitic; ss..

OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT CDS 54..758

FT /\*tag= a

FT /product= BMDSP-1

XX  
PN WO200020588-A2.

XX  
PD 13-APR-2000.

XX  
PF 01-OCT-1999; 99WO-US22908.

XX  
PR 02-OCT-1998; 98US-0165621.

XX  
PA (INCY-) INCYTE PHARM INC.

XX





protein genes, and AAG71243-AAG71319 represent the proteins they encode. AAG71320-AAG71403 represent human secreted protein fragments. The genes and their corresponding secreted proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 52 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 1061 BP; 260 A; 320 C; 241 G; 219 T; 21 other;

Query Match 91.5%; Score 880.4; DB 22; Length 1061;  
Best Local Similarity 96.1%; Pred. No. 3e-219;  
Matches 903; Conservative 8; Mismatches 23; Indels 6; Gaps 1;

29 GCTCAGTTAGACCCAGAGGGAACCATGGAAGCCCGAGCTCAGCTTCTCTCTCTCTGCT 88  
63 GCTCAGTTAGACCCAGAGGGAACCATGGAAGCCCGAGCTTCTCTCTCTCTCTGCT 122  
89 ACTCTGGTCCAGATACCCAGGGAATTTGTTGACACAGTCTCCAGCCACCTGTC 148  
123 ACTCTGGTCCAGATACCCAGGGAATTTGTTGACACAGTCTCCAGCCACCTGTC 182  
149 TTTGTCTCCAGGGGAAAGAGCCACCTCTCTGCGAGGCGAGTCAGAGTGTAGCAGCTA 208  
183 TTTGTCTCCAGGGGAAAGAGCCACCTCTCTGCGAGGCGAGTCAGAGTGTAGCAGCTA 242  
209 CTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCCGAGCTCCTCATCTATGATCATC 268  
243 CTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCCGAGCTCCTCATCTATGATCATC 302  
269 CAACAGGGCCACTGGCATCCCGAGGTTTCCAGTGGCGAGTGGTCTGGGACAGACTTCAC 328  
303 CACAGGGCCACTGGCATCCCGAGGTTTCCAGTGGCGAGTGGTCTGGGACAGACTTCAC 362  
329 TCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTATTACTGTACGAATATT 388  
363 TCTCACCATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTATTACTGTACGAATATT 422  
389 TA-----CTACTCCGTACACTTTTGGCCAGGGGACAGGCTGGAGATCAAGCAACTGT 442  
423 CAACTGGCCCTCCSANGTACACTTTTGGCCNAGGAGCAAGGTGGARATCAAGCAACTGT 482  
443 GGCTGCACCATCTGTCTTCACTTTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAAGTGC 502  
483 GGCTGCACCATCTGTCTTCACTTTTCCCGCCATCTGTATGAGCAGTTGAAATCTGGAAGTGC 542  
503 CTCTGTTGTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGT 562  
543 CTCTGTTGTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGT 602  
563 GGATAACGCCCTCCAAATCGGGTAACTCCCGAGGAGAGTGTACAGAGCAGGACAGCAAGGA 622

603 GGATAACGCCCTCCAAATCGGGTAACTCCCGAGGAGAGTGTACAGAGCAGGACAGCAAGGA 662  
623 CAGCACTACAGCCTCAGCAGCAGCCTGACGCTGAGCCTGAGCAAGCAGACTACGAGAAACACAA 682  
663 CAGCACTACAGCCTCAGCAGCAGCCTGACGCTGAGCCTGAGCAAGCAGACTACGAGAAACACAA 722  
683 AGTCTACGCTGCGAAGTCAACCATCAGGGCCTGAGCTGCGCCCGTCACAAAGAGCTTCAA 742  
723 AGTCTACGCTGCGAAGTCAACCATCAGGGCCTGAGCTGCGCCCGTCACAAAGAGCTTCAA 782  
743 CAGGGAGAGTGTAGAGGGAGAGTGTCCACCTGCTCTCTCAGTTCAGCTCAGCTGACCCC 802  
783 CAGGGAGAGTGTAGAGGGAGAGTGTCCACCTGCTCTCTCAGTTCAGCTCAGCTGACCCC 842  
803 CTCCTATCCTTTGGCCTCTGACCTTTTTCACAGGGGACCTACCCCTATTGGGGTCTCTC 862  
843 CTCCTATCCTTTGGCCTCTGACCTTTTTCACAGGGGACCTACCCCTATTGGGGTCTCTC 902  
863 CAGCTCATCTTTTCACCTCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 922  
903 CAGCTCATCTTTTCACCTCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 962  
923 GGAGAATGAATAAATAAGTGAATCTTTTGCAAAAAA 962  
963 GGAGAATGAATAAATAAGTGAATCTTTTGCAAAAAA 1002

RESULT 3  
AAS99477

ID AAS99477 standard; cDNA; 970 BP.

XX AAS99477;

XX AAS99477;

12-MAR-2002 (first entry)

Anti-human ALLIM monoclonal antibody clone Jmab-139; light chain cDNA.

Human; antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
antiallergic; antiulcer; neuroprotective; antithyroid; vasotropic;  
immunosuppressive; dermatological; antiinflammatory; hepatotropic;  
activation inducible lymphocyte immunomodulatory molecule; ALLIM;  
monoclonal antibody; allergy; rheumatoid arthritis; diabetes mellitus;  
multiple sclerosis; autoimmune thyroiditis; psoriasis; hepatitis;  
allergic contact-type dermatitis; chronic inflammatory dermatosis;  
systemic lupus erythematosus; autoimmune disorder; inflammation; ss;  
graft versus host reaction; immune rejection; intestinal immunity;  
ulcerative colitis; pneumonia; nephritis; vasculitis; pancreatitis.

Homo sapiens.

WO200187981-A2.

22-NOV-2001.

15-MAY-2001; 2001WO-JP04035.

18-MAY-2000; 2000JP-0147116.

30-MAR-2001; 2001JP-0099508.

(NISR) JAPAN TOBACCO INC.

Tsujii T, Tezuka K, Hori N;

WPI; 2002-075313/10.

P-PSDB; AAU74301.

New human monoclonal antibody that binds to activation inducible lymphocyte immunomodulatory molecule, useful for treating rheumatoid arthritis, multiple sclerosis and inflammation

Claim 45; Page 295-297; 300pp; English.

The invention relates to a novel human antibody (I), preferably a human



[illegible]









Db 601 CCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCACCCA 660  
QY 707 TCAGGGCCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGAGAA 766  
Db 661 TCAGGGCCCTGAGCTCGCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGGAGAA 720  
QY 767 GTGCCCCCACCCTGCTCCTCAGTTCAGCTCCAGCTGACCCCTCCCATCCCTTTGGCCTCTGACCC 826  
Db 721 GTGCCCCCACCCTGCTCCTCAGTTCAGCTCCAGCTGACCCCTCCCATCCCTTTGGCCTCTGACCC 780  
QY 827 TTTTTCACAGGGGACCTACCCCTATTGCGGTCCTCCAGCTCATCTTTCACCTCACCCCC 886  
Db 781 TTTTTCACAGGGGACCTACCCCTATTGCGGTCCTCCAGCTCATCTTTCACCTCACCCCC 840  
QY 887 CTCCTCCTCCTTGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAAATAAAGTGAAT 946  
Db 841 CTCCTCCTCCTTGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAAATAAAGTGAAT 900  
QY 947 CTTTGC 952  
Db 901 CTTTGC 906

RESULT 8

AAS22766

ID AAS22766 standard; cDNA; 1450 BP.

XX AAS22766;

AC AAS22766;

XX AAS22766;

DT 24-OCT-2001 (first entry)

XX Human cDNA encoding a novel human protein #332.

DE Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;

XX immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;

KW anticonvulsant; antiallergic; dermatological; haemostatic; antiasthmatic;

KW antidiabetic; immunogenic; antibody; gene therapy; neurological disorder;

KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

KW tissue regeneration; immune disorder.

XX Homo sapiens.

OS WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX P-PSDB; AAU14461.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

XX nervous system disorders, and for regenerating bone and cartilage -

XX Claim 1; Page 716; 894pp; English.

XX The invention relates to polynucleotides encoding novel human

XX proteins or their active domains. The polypeptides, polynucleotides and

XX antibodies raised against the polypeptides are used in a method of

XX treatment of a mammal and prevention of disorders caused by the aberrant

XX protein expression or activity. The polypeptides can be used as

XX molecular weight markers, food supplements, and in antibody production.

XX The polypeptides are used to identify compounds which bind to the

XX polypeptides. Polynucleotides of the invention are used as probes and

XX primers, for sequencing, for chromosome or gene mapping, in the

XX production of recombinant proteins, and in generating anti-sense DNA or

CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/elicite an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
CC The present sequence encodes a protein of the invention.  
XX

SQ Sequence 1450 BP; 332 A; 401 C; 395 G; 322 T; 0 other;

Query Match 82.4%; Score 792.8; DB 22; Length 1450;

Best Local Similarity 95.9%; Pred. No. 2.1e-196;

Matches 825; Conservative 0; Mismatches 32; Indels 3; Gaps 1;

QY 106 CCACCGGAGAAATTGTGTGACACAGCTCTCCAGCACCCCTGTCTTTGCTCCAGGGGAAA 165  
Db 591 CGACTGGCGAAATTGTGTGACGAGTCTCCAGCACCCCTGTCTTTGCTCCAGGGGAAA 650  
QY 166 GAGCCACCCCTCTCCTGAGGGGCCAGTCCAGAGTGT---AGCAGCTACTTAGCCTGGTACC 222  
Db 651 GAGCCACCCCTCTCCTGAGGGGCCAGTCCAGAGTGTGTAGCAGCAGCTACTAGCCTGGTACC 710  
QY 223 AACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGGCCACTG 282  
Db 711 AGCAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGGCCACTG 770  
QY 283 GCATCCACCCAGGTTGAGTGGCAGTGGGCTCTGGACAGACTTCACTCTCACCATCAGCA 342  
Db 771 GCATCCACAGAGTTCAGTGGCAGTGGGCTCTGGACAGACTTCACTCTCACCATCAGCA 830  
QY 343 GACTGGAGCCCGAAGATGTGGCAGCTTTATTTACTGTGCAAGTATTTTACTACTCCGTACA 402  
Db 831 GACTGGAGCCCGAAGATTTTGCAGTGTATTACTGTGCAAGTATTTTACTACTCCAGCA 890  
QY 403 CTTTGGCCAGGGACCCAGGCTGGAGATCAACAGAACTGTGGCTGCACCATCTGTCTTCA 462  
Db 891 CGTTCGGCCAAAGGACCAAAAGTGGATATCAAAACGAACTGTGGCTGCACCATCTGTCTTCA 950  
QY 463 TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTGTGCTGCTGA 522  
Db 951 TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTGTGCTGCTGA 1010  
QY 523 ATAACCTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATACAGCCCTCCCAATCGG 582  
Db 1011 ATAACCTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATACAGCCCTCCCAATCGG 1070  
QY 583 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGCCTCAGCA 642  
Db 1071 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACTACAGCCTCAGCA 1130  
QY 643 GCACCTTGACGCTGAGCAAGCAGACTACAGAAACACAAAGTCTACGCCCTGCGAAGTCA 702  
Db 1131 GCACCTTGACGCTGAGCAAGCAGACTACAGAAACACAAAGTCTACGCCCTGCGAAGTCA 1190  
QY 703 CCCATCAGGGCCCTGAGCTCGCCCGCTCAAAAGAGCTTCAACAGGGGAGAGTGTAGAGGG 762  
Db 1191 CCCATCAGGGCCCTGAGCTCGCCCGCTCAAAAGAGCTTCAACAGGGGAGAGTGTAGAGGG 1250  
QY 763 AGAAGTGCCCGCCACCTGCTCTCAGTTCAGCCTGACCCCTCCCATCTTTGGCCTCTG 822  
Db 1251 AGAAGTGCCCGCCACCTGCTCTCAGTTCAGCCTGACCCCTCCCATCTTTGGCCTCTG 1310  
QY 823 ACCCTTTTTCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTACCTCAC 882  
Db 1311 ACCCTTTTTCACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTACCTCAC 1370

QY	883	CCCCCTCCTCCTCTGGCTTAATATATGCTAATGTTGGAGAGAATGAATAATAAAGT	942	CC	anti-inflammatory diseases, nervous system disorders, and infection.
Db	1371	CCCCCTCCTCCTCCTGGCTTAATATATGCTAATGTTGGAGAGAATGAATAATAAAGT	1430	CC	The present sequence encodes a protein of the invention.
QY	943	GAATCTTTGCAAAAAA	962	XX	
Db	1431	GAATCTTTGCAAAAAA	1450	SQ	Sequence 1450 BP; 332 A; 401 C; 395 G; 322 T; 0 other;
RESULT 9				Query Match	82.4%; Score 792.8; DB 22; Length 1450;
AAS22767				Best Local Similarity	95.9%; Pred. No. 2.1e-196;
ID	AAS22767	standard; cDNA; 1450 BP.		Matches	825; Conservative 0; Mismatches 32; Indels 1;
AC	AAS22767;				
XX	24-OCT-2001	(first entry)			
DT	Human	cDNA encoding a novel human protein #333.			
DE	Human;	novel protein; ss; Antianaemic; osteopathic; antiinflammatory;			
XX	immunomodulatory;	cytostatic; neuroprotective; vulnery; nootropic;			
KW	anticonvulsant;	antiarthritic; cerebroprotective; antifungal; antiviral;			
KW	antibacterial;	antiallergic; dermatological; haemostatic; antiasthmatic;			
KW	thrombolytic;	immunogen; antibody; gene therapy; neurological disorder;			
KW	Parkinson's disease;	inflammatory disorder; cancer; asthma; osteoporosis;			
KW	tissue regeneration;	immune disorder.			
OS	Homo sapiens.				
XX	WO200155437-A2.				
PN	02-AUG-2001.				
PD	25-JAN-2001;	2001WO-US02623.			
XX	25-JAN-2000;	2000US-0491404.			
PR	(HYSE-) HYSEQ INC.				
XX	Tang YT, Liu C, Drmanac RT;				
PI	WPI; 2001-451939/48.				
XX	P-PSDB; AAU14462.				
DR	Isolated polypeptides useful for treating anti-inflammatory diseases,				
XX	nervous system disorders, and for regenerating bone and cartilage -				
PT	Claim 1; Page 716-717; 894pp; English.				
PS	The invention relates to polynucleotides encoding novel human				
CC	proteins or their active domains. The polypeptides, polynucleotides and				
CC	antibodies raised against the polypeptides are used in a method of				
CC	treatment of a mammal and prevention of disorders caused by the aberrant				
CC	protein expression or activity. The polypeptides can be used as				
CC	molecular weight markers, food supplements, and in antibody production.				
CC	The polypeptides are used to identify compounds which bind to the				
CC	polypeptides. Polynucleotides of the invention are used as probes and				
CC	primers, for sequencing, for chromosome or gene mapping, in the				
CC	production of recombinant proteins, and in generating anti-sense DNA or				
CC	RNA and in gene therapy. Polypeptides of the invention can be used to				
CC	target drugs to a tumour, in assays to determine biological activity, to				
CC	raise antibodies/elicite an immune response, to determine quantitative				
CC	protein levels, as tissue markers, and to isolate receptors or ligands.				
CC	Polypeptides of the invention may also be useful in treating platelet				
CC	disorders, stem cell disorders, regenerating bone, cartilage, tendon,				
CC	ligament and/or nerve tissue, wound healing, treating burns, promoting				
CC	the proliferation, differentiation and survival of stem cells, as a				
CC	contraceptive, treating osteoporosis and osteoarthritis, anaemia,				
CC	Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral				
CC	sclerosis, stroke, immune deficiencies resulting from bacterial, viral or				
CC	fungal infection or from autoimmunity, cancer, allergy, asthma,				
CC	graft-versus-host disease, eczema, haemophilia, thrombosis,				



XX AAS22768;  
AC  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human CDNA encoding a novel human protein #334.  
XX  
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;  
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200155437-A2.  
PD  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02623.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Llu C, Drmanac RT;  
XX  
DR WPI; 2001-451939/48.  
DR P-PSDB; AAU14463.  
XX  
PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
PT nervous system disorders, and for regenerating bone and cartilage -  
XX  
PS Claim 1; Page 717; 894pp; English.  
XX  
CC The invention relates to polynucleotides encoding novel human  
CC proteins or their active domains. The polypeptides, polynucleotides and  
CC antibodies raised against the polypeptides are used in a method of  
CC treatment of a mammal and prevention of disorders caused by the aberrant  
CC protein expression or activity. The polypeptides can be used as  
CC molecular weight markers, food supplements, and in antibody production.  
CC The polypeptides are used to identify compounds which bind to the  
CC polypeptides. Polynucleotides of the invention are used as probes and  
CC primers, for sequencing, for chromosome or gene mapping, in the  
CC production of recombinant proteins, and in generating anti-sense DNA or  
CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/ elicit an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
CC The present sequence encodes a protein of the invention.  
XX  
SQ Sequence 1450 BP; 332 A; 401 C; 395 G; 322 T; 0 other;

Query Match 82.4%; Score 792.8; DB 22; Length 1450;  
Best Local Similarity 95.9%; Pred. No. 2.le-196;  
Matches 825; Conservative 0; Mismatches 32; Indels 3; Gaps 1;  
QY 106 CCACCGGAGAAATGTGTTGACACAGTCTCCAGCCACCTGTCTTTGTCTCCAGGGGAAA 165  
Db 591 CGACTGGCGAAATGTGTTGACCGAGTCTCCAGGCACCTGTCTTTGTCTCCAGGGGAAA 650

QY 166 GAGCCACCCCTCTCTGTCAGGGCCAGTCAGAGTGT---AGCAGTACTTAGCCTGGTACC 222  
Db 651 GAGCCACCCCTCTCTGTCAGGGCCAGTCAGAGTGT---AGCAGTACTTAGCCTGGTACC 710  
QY 223 AACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCAAACAGGGCCACTG 282  
Db 711 AGCAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCAGGCGCCACTG 770  
QY 283 GCATCCACCCAGGTTGAGTGGCAGTGGTCTGGGACAGACTTCACTCTCACCATCAGCA 342  
Db 771 GCATCCACAGAGGTTGAGTGGCAGTGGTCTGGGACAGACTTCACTCTCACCATCAGCA 830  
QY 343 GACTGGAGCCCGAAGATGTGGCACTTTATTACTGTGAGCAATATTTTACTACTCCGTACA 402  
Db 831 GACTGGAGCCCTGAAGATTTTGCAGTGTATTACTGTGAGCAATATTTTACTACTCCGTACA 890  
QY 403 CTTTGGCCAGGGGACCCAGGCTGGAGATCAAAACGAACTGTGGCTGCACCATCTGTCTTCA 462  
Db 891 CGTTCGGCCAAAGGACCAAGTGGATATCAAAACGAACTGTGGCTGCACCATCTGTCTTCA 950  
QY 463 TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCTGCTCTGTGTGCTGCTGTA 522  
Db 951 TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAATCTGCTCTGTGTGCTGCTGTA 1010  
QY 523 ATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATAACGCCCTCCAAATCGG 582  
Db 1011 ATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAGGTGGATAACGCCCTCCAAATCGG 1070  
QY 583 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACACCTCAGCA 642  
Db 1071 GTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACACCTCAGCA 1130  
QY 643 GCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCA 702  
Db 1131 GCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCCTGCGAAGTCA 1190  
QY 703 CCCATCAGGGCCTGAGCTGCGCCCTGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGG 762  
Db 1191 CCCATCAGGGCCTGAGCTGCGCCCTGTCACAAAGAGCTTCAACAGGGGAGAGTGTAGAGGG 1250  
QY 763 AGAAGTGCCCCCACCCTGCTCCTCAGTCCAGCTCCACCCCTCCCATCTTGGCCTCTG 822  
Db 1251 AGAAGTGCCCCCACCCTGCTCCTCAGTCCAGCTCCACCCCTCCCATCTTGGCCTCTG 1310  
QY 823 ACCCTTTTCCACAGGGGACCTACCCCTATTTGGGGTCTCCAGCTCATCTTTACCTCAC 882  
Db 1311 ACCCTTTTCCACAGGGGACCTACCCCTATTTGGGGTCTCCAGCTCATCTTTACCTCAC 1370  
QY 883 CCCCCTCCTCCTCTGGGCTTTAATTTATGCTAATTTGGAGGAGAGTAATAAAGT 942  
Db 1371 CCCCCTCCTCCTCCTCTGGGCTTTAATTTATGCTAATTTGGAGGAGAGTAATAAAGT 1430  
QY 943 GAATCTTTGCAAAAAA 962  
Db 1431 GAATCTTTGCAAAAAA 1450  
RESULT 11  
AAS22769  
ID AAS22769 standard; cDNA; 1450 BP.  
XX  
AC AAS22769;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human CDNA encoding a novel human protein #335.  
XX  
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;  
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;



tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02623.

25-JAN-2000; 2000US-0491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

P-PSDB; AAU14464.

Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage -

Claim 1; Page 717-718; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/ elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting the proliferation, differentiation and survival of stem cells, as a contraceptive, treating osteoporosis and osteoarthritis, anaemia, Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or fungal infection or from autoimmunity, cancer, allergy, asthma, graft-versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory diseases, nervous system disorders, and infection. The present sequence encodes a protein of the invention.

Sequence 1450 BP; 332 A; 401 C; 395 G; 322 T; 0 other;

Query Match		82.4%;	Score 792.8;	DB 22;	Length 1450;
Best Local Similarity		95.9%;	Pred. No. 2.1e-196;		
Matches		825;	Conservative	0;	Mismatches 32; Indels 3; Gaps 1;
QY	106	CCACCGGAGAAATTGTTGACACAGTCTCCAGCCACCTGTCTTTGTCTCCAGGGGAAA	165		
Db	591	CGACTGGCGAAATTGTTGACGAGTCTCCAGGCACCTGTCTTTGTCTCCAGGGGAAA	650		
QY	166	GAGCCACCTCTCCTGCAGGSCCAGTCAGAGTGT---AGCAGCTACTTAGCCTGGTACC	222		
Db	651	GAGCCACCTCTCCTGCAGGSCCAGTCAGAGTGTTAGCAGCAGCTACTTAGCCTGGTACC	710		
QY	223	AACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGCCACTG	282		
Db	711	AGCAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGGTGCATCCAGGGCCACTG	770		
QY	283	GCATCCACCCAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACCTCACCATCAGCA	342		
Db	771	GCATCCACAGACAGGTTTCAGTGGCAGTGGGTCTGGGACAGACTTCACCTCACCATCAGCA	830		
QY	343	GACTGGAGCCCGAAGATGTGGCACTTTATTACTGTTCAGCAATATTTTACTACTCCGTACA	402		

Db	831	
----	-----	--

FT misc\_RNA 728..1044  
FT /tag= d  
FT /note= "constant region"  
FT polyA\_signal 1221..1226  
FT /tag= e  
PN WO9106305-A.  
XX  
PD 16-MAY-1991.  
XX  
PF 06-NOV-1990; 90WO-US06426.  
XX  
PR 07-NOV-1989; 89US-0432700.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIB.  
XX  
PI Shuford WW, Harris LJ, Raff HV;  
XX  
DR WPI; 1991-163947/22.  
DR P-PSDB; AAR12128.  
XX  
PT Oligomeric immunoglobulin(s) with high avidity for antigen(s) -  
PT formed by duplicating esp. variable region of light chain of IgG  
PT class  
XX  
PS Claim 42; Fig 16; 104pp; English.  
XX  
CC This sequence encodes an aberrant light chain containing two L'V  
CC regions. Antibody molecules of the invention can include one or two  
CC such aberrant light chains to produce heavier antibodies. These  
CC heavier antibodies were found to have higher avidity than antibodies  
CC with just a single copy of the L'V region. The clone is not  
CC complete; it starts with the G of the ATG initiator codon.  
CC See also AAQ11879 and AAQ11880.  
XX  
SQ Sequence 1242 BP; 291 A; 383 C; 302 G; 266 T; 0 other;  
  
Query Match 82.4%; Score 792.4; DB 12; Length 1242;  
Best Local Similarity 95.7%; Pred. No. 2.6e-196;  
Matches 827; Conservative 0; Mismatches 31; Indels 6; Gaps 1;  
  
QY 95 GCTCCAGATACACCGGAGAAATTGTGTGACACAGTCTCAGCCACCCCTGTCTTGTG 154  
DB 379 GATCAAAACATACACCGGAGAAATTGTGTGACACAGTCTCAGCCACCCCTGTCTTGTG 438  
  
QY 155 TCCAGGGGAAGAGCCACCCTCTCTCGAGGGCCAGTCTAGCAGCTACTTAGC 214  
DB 439 TCCAGGGGAAGAGCCACCCTCTCTCGAGGGCCAGTCTAGCAGCTACTTAGC 498  
  
QY 215 CTGGTACCAACAGAAACCTGCGCCAGGCTCCAGGCTCTCATCTATGATGATCCACAG 274  
DB 499 CTGGTACCAACAGAAACCTGCGCCAGGCTCCAGGCTCCAGGCTCTCATCTATGATGATGATCCACAG 558  
  
QY 275 GGCCACTGGCATCCCAACCCAGGTTTCACTGGCAGTGGGTCTGGACAGACTTCACTCTCAC 334  
DB 559 GGCCACTGGCATCCCAACCCAGGTTTCACTGGCAGTGGGTCTGGACAGACTTCACTCTCAC 618  
  
QY 335 CATCAGCAGACTGGAGCCCGGAGATGTGGCACTTTATTACTGTGAGC-----AATATTT 388  
DB 619 CATCAGCAGCTAGAGCCTGAAGATTTTGCAGTTTATTACTGTCAACACCCGTGACAATTG 678  
  
QY 389 TACTACTCCGTACACTTTTGGCCAGGGGACAGGCTGGAGATCAAAACGAACTGTGGCTGC 448  
DB 679 GCCTCCGGGGCCACTTTTCGGCGGAGGGACCAAGTGGAGATCAAAACGAACTGTGGCTGC 738  
  
QY 449 ACCATCTGTCTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCTCTGT 508  
DB 739 ACCATCTGTCTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCTCTGT 798  
  
QY 509 TGTGTGCTCTGTAATAACTTCTATCCAGAGAGGGCCAAAGTACAGTGGAGGTGATAA 568  
DB 799 TGTGTGCTCTGTAATAACTTCTATCCAGAGAGGGCCAAAGTACAGTGGAGGTGATAA 858

QY 569 CGCCCTCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAGGACAGCAC 628  
DB 859 CGCCCTCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAGGACAGCAC 918  
  
QY 629 CTACAGCCTCAGCAGCAGCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTA 688  
DB 919 CTACAGCCTCAGCAGCAGCCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTA 978  
  
QY 689 CGCCTGCGAAGTCAACCATCAGGGCCTGAGCTCGCCCGTCAACAAAGAGCTTCAACAGGG 748  
DB 979 CGCCTGCGAAGTCAACCATCAGGGCCTGAGCTCGCCCGTCAACAAAGAGCTTCAACAGGG 1038  
  
QY 749 AGAGTGTAGAGGGAGAGTGCCTCCACCTGCTCTCAGTTCCAGCCTGACCCCTCCCA 808  
DB 1039 AGAGTGTAGAGGGAGAGTGCCTCCACCTGCTCTCAGTTCCAGCCTGACCCCTCCCA 1098  
  
QY 809 TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTC 868  
DB 1099 TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTC 1158  
  
QY 869 ATCTTTACCTCACCCCT 928  
DB 1159 ATCTTTACCTCACCCCT 1218  
  
QY 929 TGAATAAATAAAGTGAATCTTTGC 952  
DB 1219 TGAATAAATAAAGTGAATCTTTGC 1242  
  
RESULT 13  
ABK64543  
ID ABK64543 standard; DNA; 1244 BP.  
XX  
AC ABK64543;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human benign prostatic hyperplasia gene #438.  
XX  
DE Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200212440-A2.  
XX  
PD 14-FEB-2002.  
XX  
PF 07-AUG-2001; 2001WO-US24708.  
XX  
PR 07-AUG-2000; 2000US-223323P.  
PR 05-JUN-2001; 2001US-0873319.  
XX  
PA (GENE-) GENE LOGIC INC.  
PA (NISR ) JAPAN TOBACCO INC.  
XX  
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
XX  
DR WPI; 2002-257476/30.  
XX  
PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
PT detecting expression levels of one or more genes in prostate cells from  
PT patient that are differentially regulated compared to normal prostate  
PT cells -  
XX  
PS Disclosure; Page 256-257; 444pp; English.  
XX  
CC The invention relates to a method of diagnosing (I) the onset or  
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
CC or identifying an agent that modulates the onset or progression of BPH.  
CC The method is based on changes in gene expression in BPH tissue isolated  
CC from patients exhibiting different clinical states of prostate  
CC hyperplasia as compared to normal prostate tissue. (I) comprises  
CC detecting the expression levels of one or more genes in prostate cells

CC from the subject that are differentially regulated compared to normal  
CC prostate cells. (II) comprises preparing a first gene expression profile  
CC of BPH cells or BPH-like cell population, exposing the cells to the  
CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles.  
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is  
CC useful for identifying an agent that modulates the onset or progression  
CC of BPH. The methods are useful to present information identifying  
CC the expression level in a tissue or cells, by comparing the expression  
CC level of genes given in the specification in the tissue or cells to the  
CC level of expression of gene in the database, and displaying the  
CC expression levels of at least one gene in the tissue or cell sample  
CC compared to the expression level in BPH. Agents using (II) are useful for  
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human  
CC benign prostatic hyperplasia gene sequences of the invention.

XX	Sequence 1244 BP; 292 A; 383 C; 302 G; 267 T; 0 other;
SQ	
Query Match 82.4%; Score 792.4; DB 24; Length 1244;	
Best Local Similarity 95.7%; Pred. No. 2.6e-196;	
Matches 827; Conservative 0; Mismatches 31; Indels 6; Gaps 1;	
QY	95 GCTCCAGATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGTCTTGTGTC 154
Db	
QY	381 GATCAACATACACCGGAGAAATTGTGTGACACAGTCTCCAGCCACCCCTGTCTTGTGTC 440
Db	
QY	155 TCCAGGGGAAAGAGCCACCCCTCTCTGCGAGGGCCAGTCAAGAGTGTAGCAGCTACTAGC 214
Db	
QY	441 TCCAGGGGAAAGAGCCACCCCTCTCTGCGAGGGCCAGTCAAGAGTGTAGCAGCTACTAGC 500
Db	
QY	215 CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGATCAACACAG 274
Db	
QY	501 CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGATCAACACAG 560
Db	
QY	275 GGCACACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGCTGGACAGACTTCACCTCTAC 334
Db	
QY	561 GGCACACTGGCATCCACCCAGGTTTCAGTGGCAGTGGGCTGGACAGACTTCACCTCTAC 620
Db	
QY	335 CATCAGCAGACTGGAGCCCGAAGATGTGGCACTTTATTACTGTGACG-----AATATT 388
Db	
QY	621 CATCAGCAGCTAGAGCCTGAAGATTTTGCAGTTTATTACTGTCAACACCCGTGACAATTG 680
Db	
QY	389 TACTACTCCGTACACTTTTGGCCAGGGGACCAAGGCTGGAGATCAACAGAACTGTGGCTGC 448
Db	
QY	681 GCCTCCGGGGCCACTTTCGGCGGAGGGACCAAGGTGGAGATCAACAGAACTGTGGCTGC 740
Db	
QY	449 ACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCTCTGT 508
Db	
QY	741 ACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCTCTGT 800
Db	
QY	509 TGTGTGCTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATAA 568
Db	
QY	801 TGTGTGCTGCTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATAA 860
Db	
QY	569 CGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC 628
Db	
QY	861 CGCCCTCCAAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCAC 920
Db	
QY	629 CTACAGCCTCAGCAGCACCCTGACGCTGAGCAACAGCAGACTACGAGAAACACAAAGTCTA 688
Db	
QY	921 CTACAGCCTCAGCAGCACCCTGACGCTGAGCAACAGCAGACTACGAGAAACACAAAGTCTA 980
Db	
QY	689 CGCCTGCGAAGTCAACCCATCAGGCGCTGAGCTCGCCGTCAACAAAGAGCTTCAACAGGGG 748
Db	
QY	981 CGCCTGCGAAGTCAACCCATCAGGCGCTGAGCTCGCCGTCAACAAAGAGCTTCAACAGGGG 1040
Db	
QY	749 AGAGTGTAGAGGGAGAGTGGCCCGCCACCTGCTCTCAGTTCAGCCTGACCCCTCCCA 808
Db	
QY	1041 AGAGTGTAGAGGGAGAGTGGCCCGCCACCTGCTCTCAGTTCAGCCTGACCCCTCCCA 1100
Db	
QY	809 TCCTTTGGCCTCTGACCCCTTTTTCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTC 868
Db	
QY	1101 TCCTTTGGCCTCTGACCCCTTTTTCACAGGGGACCTACCCCTATTGCGGTCTCTCCAGCTC 1160
Db	

QY	869 ATCTTTACCTCACCCTCCCTCCTCCTCCTTAAATTTATGCTTAATGTTGGAGAGAA 928
Db	
QY	1161 ATCTTTACCTCACCCTCCCTCCTCCTCCTTAAATTTATGCTTAATGTTGGAGAGAA 1220
Db	
QY	929 TGAATAAATAAAGTGAATCTTTGC 952
Db	
QY	1221 TGAATAAATAAAGTGAATCTTTGC 1244
Db	
RESULT 14	
ABL65461	
ID	ABL65461 standard; DNA; 1244 BP.
XX	AC ABL65461;
XX	15-MAY-2002 (first entry)
XX	Lung cancer related gene sequence SEQ ID NO:3798.
DE	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW	gene; ds.
XX	Homo sapiens.
OS	WO200194629-A2.
XX	13-DEC-2001.
PN	30-MAY-2001; 2001WO-US10838.
XX	05-JUN-2000; 2000US-209473P.
PR	05-JUN-2000; 2000US-209531P.
PR	18-SEP-2000; 2000US-233133P.
PR	18-SEP-2000; 2000US-233617P.
PR	20-SEP-2000; 2000US-234009P.
PR	20-SEP-2000; 2000US-234034P.
PR	20-SEP-2000; 2000US-234052P.
PR	22-SEP-2000; 2000US-234509P.
PR	22-SEP-2000; 2000US-234567P.
PR	25-SEP-2000; 2000US-234923P.
PR	25-SEP-2000; 2000US-234924P.
PR	25-SEP-2000; 2000US-235077P.
PR	25-SEP-2000; 2000US-235082P.
PR	25-SEP-2000; 2000US-235134P.
PR	25-SEP-2000; 2000US-235280P.
PR	26-SEP-2000; 2000US-235637P.
PR	26-SEP-2000; 2000US-235638P.
PR	27-SEP-2000; 2000US-235711P.
PR	27-SEP-2000; 2000US-235720P.
PR	27-SEP-2000; 2000US-235840P.
PR	27-SEP-2000; 2000US-235863P.
PR	28-SEP-2000; 2000US-236028P.
PR	28-SEP-2000; 2000US-236032P.
PR	28-SEP-2000; 2000US-236033P.
PR	28-SEP-2000; 2000US-236034P.
PR	28-SEP-2000; 2000US-236109P.
PR	28-SEP-2000; 2000US-236111P.
PR	29-SEP-2000; 2000US-236842P.
PR	29-SEP-2000; 2000US-236891P.
PR	02-OCT-2000; 2000US-237172P.
PR	02-OCT-2000; 2000US-237173P.
PR	02-OCT-2000; 2000US-237278P.
PR	02-OCT-2000; 2000US-237294P.
PR	02-OCT-2000; 2000US-237295P.
PR	02-OCT-2000; 2000US-237316P.
PR	03-OCT-2000; 2000US-237425P.
PR	03-OCT-2000; 2000US-237598P.
PR	03-OCT-2000; 2000US-237604P.
PR	03-OCT-2000; 2000US-237606P.
PR	03-OCT-2000; 2000US-237608P.



PR 01-NOV-2000; 2000US-244867P.  
XX 01-NOV-2000; 2000US-245084P.  
PA (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
DR Screening for anti-neoplastic agent involves exposing cells to a  
XX chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
PT Claim 1; SEQ ID 3798; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 1244 BP; 292 A; 383 C; 302 G; 267 T; 0 other;

Query Match  
Best Local Similarity 82.4%; Score 792.4; DB 24; Length 1244;  
Matches 827; Conservative 0; Mismatches 31; Indels 6; Gaps 1;

QY 95 GCTCCAGATACACCGAGAAATTGTTGACAGTCTCCAGCCACCTGCTTTGTC 154  
Db 381 GATCAACATACACCGAGAAATTGTTGACAGTCTCCAGCCACCTGCTTTGTC 440  
QY 155 TCCAGGGGAAAGAGCCACCTCTCTGCGAGGCGAGTGTAGCAGCTACTTACC 214  
Db 441 TCCAGGGGAAAGAGCCACCTCTCTGCGAGGCGAGTGTAGCAGCTACTTACC 500  
QY 215 CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCTCATCTATGATGATCAACAG 274  
Db 501 CTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCTCATCTATGATGATCAACAG 560  
QY 275 GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGGACAGACTTCACTCTAC 334  
Db 561 GGCCACTGGCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGGACAGACTTCACTCTAC 620  
QY 335 CATCAGCAGACTGGAGCCCGAAGATGCGCACTTTATTACTGTGAGC-----AATATT 388  
Db 621 CATCAGCAGCTAGAGCCCTGAAGATTGTCAGTTTATTACTGTCAACACCCGTGACAATTG 680  
QY 389 TACTACTCCGTACACTTTTGGCCAGGACAGGCTGGAGTCAACAGAACTGTGGCTGC 448  
Db 681 GCCTCCGGGGCCACTTTGCGGAGGAGGACCAAGTGGAGATCAACAGAACTGTGGCTGC 740  
QY 449 ACCATCTGCTTTCATCTTCCCGCATCTGATGAGCAGTTGAAATCTGGAACCTGCTGT 508  
Db 741 ACCATCTGCTTTCATCTTCCCGCATCTGATGAGCAGTTGAAATCTGGAACCTGCTGT 800  
QY 509 TGTGTGCTGCTGAATACTTCTATCCAGAGAGGCGCAAGTACAGTGAAGTGGATAA 568  
Db 801 TGTGTGCTGCTGAATACTTCTATCCAGAGAGGCGCAAGTACAGTGAAGTGGATAA 860  
QY 569 CGCCCTCCCAATCGGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGCAAGSACAGCAG 628

Db 861 CGCCCTCCAATCGGTAACCTCCAGGAGAGTGTACAGAGCAGGACAGGACAGCAGCAC 920  
QY 629 CTACAGCCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 688  
Db 921 CTACAGCCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 980  
QY 689 CGCCTGCGAAGTCAACCATCAGGCGCTGAGCTGCGCCCGTCAACAGAGCTTCAACAGGGG 748  
Db 981 CGCCTGCGAAGTCAACCATCAGGCGCTGAGCTGCGCCCGTCAACAGAGCTTCAACAGGGG 1040  
QY 749 AGAGTGTAGAGGAGAGAGTGCCTCCACCTGCTCTCAGTTCAGCTGACCCCTCCCA 808  
Db 1041 AGAGTGTAGAGGAGAGAGTGCCTCCACCTGCTCTCAGTTCAGCTGACCCCTCCCA 1100  
QY 809 TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTCGGGTCTCCAGCTC 868  
Db 1101 TCCTTTGGCCTCTGACCCCTTTTCCACAGGGGACCTACCCCTATTCGGGTCTCCAGCTC 1160  
QY 869 ATCTTTACCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 928  
Db 1161 ATCTTTACCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1220  
QY 929 TGAATAAATAAAGTGAATCTTTC 952  
Db 1221 TGAATAAATAAAGTGAATCTTTC 1244

RESULT 15

AAS22530

ID AAS22530 standard; cDNA; 1721 BP.

XX AAS22530;

AC AAS22530;

XX 24-OCT-2001 (first entry)

XX Human cDNA encoding a novel human protein #96.

DE Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;

XX immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;

XX anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

XX antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

XX parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

XX tissue regeneration; immune disorder.

XX Homo sapiens.

OS WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX P-PSDB; AAU14225.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

XX nervous system disorders, and for regenerating bone and cartilage -

XX Claim 1; Page 301-302; 894pp; English.

XX The invention relates to polynucleotides encoding novel human

XX proteins or their active domains. The polypeptides, polynucleotides and

XX antibodies raised against the polypeptides are used in a method of

XX treatment of a mammal and prevention of disorders caused by the aberrant

XX protein expression or activity. The polypeptides can be used as



CC molecular weight markers, food supplements, and in antibody production.  
CC The polypeptides are used to identify compounds which bind to the  
CC polypeptides. Polynucleotides of the invention are used as probes and  
CC primers, for sequencing, for chromosome or gene mapping, in the  
CC production of recombinant proteins, and in generating anti-sense DNA or  
CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/elicit an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
CC ligament and/or nerve tissue; wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
CC The present sequence encodes a protein of the invention.

XX  
SQ Sequence 1721 BP; 394 A; 473 C; 428 G; 426 T; 0 other;

Query Match		80.2%;	Score 771.2;	DB 22;	Length 1721;
Best Local Similarity		95.7%;	Pred. NO. 9.4e-191;		
Matches 815;		Conservative 0;	Mismatches 33;	Indels 4;	Gaps 2;
QY	106	CCACCGGAGAAATTGTTGACACAGTCTCCAGCCACCCCTGTCCTTGTCTCCAGGGGAAA	165		
Db	601	CGACTGGCGAAATTGTTGACGAGTCTCCAGGCACCCCTGTCCTTGTCTCCAGGGGAAA	660		
QY	166	GAGCCACCTCTCCTGCGAGGCGCAGTCAGAGTGT---AGCAGTACTTAGCCTGGTACC	222		
Db	661	GAGCCACCTCTCCTGCGAGGCGCAGTCAGAGTGTAGCAGCAGCTACTTAGCCTGGTACC	720		
QY	223	AACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGCCACTG	282		
Db	721	AGCAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGGGCCACTG	780		
QY	283	GCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGACAGACTTCACCTCACCATCAGCA	342		
Db	781	GCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGACAGACTTCACCTCACCATCAGCA	840		
QY	343	GACTGGAGCCGCAAGATGTGGCACTTTATTACTGTGACGAATATTTTACTACTCCGTACA	402		
Db	841	GACTGGAGCCGCAAGATTTTGCAGTGTATTACTGTGACGAGCGGGTCGTATTCGCCCGA	900		
QY	403	CTTTTGGCCAGGGACCGCTGGAGATCAAACGAACCTGTGGCTGCACCATCTGTCTTCA	462		
Db	901	CGTTCCGCCAAGGACCAAGGTGGAATCAAACGAACCTGTGGCTGCACCATCTGTCTTCA	960		
QY	463	TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCCTCTGTTGTGCTGCTGA	522		
Db	961	TCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCCTCTGTTGTGCTGCTGA	1020		
QY	523	ATAACTTCTATCCACAGAGGCGCAAGTACAGTGGAGGTGGATAACGCCCT-CCAAATCG	581		
Db	1021	ATAACTTCTATCCACAGAGGCGCAAGTACAGTGGAGGTGGATAACGCCCTCCCAATCG	1080		
QY	582	GGTAACCTCCAGGAGTGTACAGAGCAGGACAGCAAGGACAGCACCCTACAGCCTCAGC	641		
Db	1081	GGTAACCTCCAGGAGTGTACAGAGCAGGACAGCAAGGACAGCACCCTACAGCCTCAGC	1140		
QY	642	AGCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTC	701		
Db	1141	AGCACCTGACGCTGAGCAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTC	1200		
QY	702	ACCATCAGGGCCTGAGCTGCGCCCGTCAACAAGAGCTTCAACAGGGGAGAGTGTAGAGG	761		
Db	1201	ACCATCAGGGCCTGAGCTGCGCCCGTCAACAAGAGCTTCAACAGGGGAGAGTGTAGAGG	1260		
QY	762	GAGAGTGGCCCCACCTGCTCCTCAGTTCAGCCTCCACCTCCCTTGGCCTCT	821		

Db	1261	GAGAAAGTGCCCCACCTGCTCCTCAGTTCAGCTCCAGCCTGACCCCTCCCATCCTTTGGCCTCT	1320
QY	822	GACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCCCTCCAGCTCATCTTTACCTCA	881
Db	1321	GACCCCTTTTCCACAGGGGACCTACCCCTATTGCGGTCCCTCCAGCTCATCTTTACCTCA	1380
QY	882	CCCCCTCCTCCTTGGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAATAAG	941
Db	1381	CCCCCTCCTCCTTGGCTTTAATTATGCTAATGTTGGAGGAGAAATGAATAATAAG	1440
QY	942	TGAATCTTTGCA	953
Db	1441	TGAATCTTTGCA	1452

Search completed: July 25, 2003, 15:52:49  
Job time : 227.852 secs









**RESULT** 11  
AC090491/c  
**LOCUS**  
DEFINITION AC090491 bp DNA linear HTG 24-FEB-2001  
Mus musculus clone RP23-12K1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 43  
unordered pieces.  
AC090491  
AC090491

AC090491.1 GI:13122740  
HTG; HTGS\_PHASE1.  
Mus musculus.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
1 (bases 1 to 204019)  
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Baliya,V.,  
Dedhia,N.N., de la Bastide,M., Huang,E.N., King,L., Kirchoff,K.A.,  
Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,  
Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Vil,M.D.  
and Zutavern,T.  
Mouse Genomic Sequence  
Unpublished  
2 (bases 1 to 204019)  
McCombie,W.R.  
Direct Submission  
Submitted (24-FEB-2001) Lita Annenberg Hazen Genome Sequencing  
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA

12288:	contig of 12288 bp in length
12369:	gap of unknown length
24613:	contig of 12244 bp in length
24694:	gap of unknown length
24695:	contig of 11823 bp in length
36517:	gap of unknown length
36598:	contig of 9121 bp in length
45719:	gap of unknown length
45800:	contig of 8540 bp in length
54340:	gap of unknown length
54341:	gap of unknown length
54421:	contig of 8463 bp in length
62884:	gap of unknown length
62965:	contig of 7987 bp in length
70952:	gap of unknown length
71033:	contig of 6655 bp in length
77688:	gap of unknown length
77769:	contig of 6628 bp in length
84398:	gap of unknown length
84478:	contig of 6333 bp in length
90811:	gap of unknown length
90892:	contig of 5922 bp in length
96814:	gap of unknown length
96895:	contig of 5229 bp in length
102124:	gap of unknown length
102205:	contig of 5070 bp in length
107275:	gap of unknown length
107356:	contig of 4943 bp in length
112299:	gap of unknown length
112380:	contig of 4874 bp in length
117254:	gap of unknown length
117335:	contig of 4872 bp in length
122207:	gap of unknown length
122288:	contig of 4575 bp in length
126863:	gap of unknown length
126944:	contig of 4125 bp in length
131069:	contig of 4125 bp in length



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 14:13:55 ; Search time 6693.77 Seconds  
(without alignments)  
11069.353 Million cell updates/sec

Title: US-09-806-276A-4  
Perfect score: 2546  
Sequence: 1 cggggctagccggagagacc.....cagctanaacgagggcagta 2546

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2370.8	93.1	188359	9	AC006369	AC006369 Homo sapi
2	2165.6	85.1	2768	9	HSM801810	AL136842 Homo sapi
3	1840	72.3	1892	9	BC019270	BC019270 Homo sapi
4	1215.8	47.8	93663	9	AC022409	AC022409 Homo sapi
5	1051.4	41.3	1118	9	AF164118	AF164118 Homo sapi
6	925.8	36.4	941	9	AF104857	AF104857 Homo sapi
7	904.6	35.5	216905	2	AC091332	AC091332 Mus muscu
8	895.6	35.2	292420	2	AC055705	AC055705 Mus muscu
9	769.8	30.2	1481	10	BC034714	BC034714 Mus muscu
10	761.8	29.9	765	9	AF094521	AF094521 Homo sapi
11	747.2	29.3	204019	2	AC090491	AC090491 Mus muscu
12	692.8	27.2	985	10	BC021409	BC021409 Mus muscu
13	393	15.4	432	11	G23498	G23498 human STS W
14	117.4	4.6	741	9	AF098290	AF098290 Homo sapi
15	117.4	4.6	1670	9	BC022337	BC022337 Homo sapi
16	117.4	4.6	1976	9	AF163840	AF163840 Homo sapi
17	117.4	4.6	1976	9	HSFA001436	AF001436 Human clo
18	117.4	4.6	2128	9	AK098292	AK098292 Homo sapi
19	117.4	4.6	78242	2	AP000483	AP000483 Homo sapi
20	117.4	4.6	171304	2	AC090297	AC090297 Homo sapi
21	117.4	4.6	218230	2	AC090390	AC090390 Homo sapi
22	117.4	4.6	233089	9	AC000353	AC000353 Homo sapi
23	115.8	4.5	1644	9	BC005406	BC005406 Homo sapi
24	115.2	4.5	149394	2	AP000828	AP000828 Homo sapi
25	108.4	4.3	1466	10	BC034884	BC034884 Mus muscu
26	108.4	4.3	203464	2	AC127271	AC127271 Mus muscu
27	107.6	4.2	51198	2	AP001279	AP001279 Homo sapi
28	88.8	3.5	200498	2	AP000944	AP000944 Homo sapi
29	64.4	2.5	110000	2	CEY111B2_0	Z98857 Caenorhabdi
30	64.4	2.5	257703	3	CEY111B2A	AL132904 Caenorhab
31	63.8	2.5	30726	2	AC117269	AC117269 Dictyoste
32	62.8	2.5	3364	14	TSP418778	AJ418778 Tomato sp
33	62.4	2.5	349980	6	AX344558	AX344558 Sequence
34	62	2.4	11812	6	AX251500	AX251500 Sequence
35	62	2.4	11812	6	AX278047	AX278047 Sequence
36	62	2.4	11812	6	AX323818	AX323818 Sequence
37	62	2.4	11812	6	AX347021	AX347021 Sequence
38	61.8	2.4	163547	2	AC111523	AC111523 Rattus no
39	61	2.4	131682	9	AL672277	AL672277 Human DNA
40	59.6	2.3	349980	6	AX344557	AX344557 Sequence
41	59.2	2.3	56772	3	AC006712	AC006712 Caenorhab
42	59	2.3	8079	6	AX356488	AX356488 Sequence
43	58.6	2.3	92971	9	AC023165	AC023165 Homo sapi
44	58.6	2.3	161523	2	AC084039	AC084039 Homo sapi
45	58.6	2.3	294914	3	AE003603	AE003603 Drosophila

ALIGNMENTS

RESULT 1  
AC006369  
LOCUS AC006369 188359 bp DNA linear PRI 30-SEP-2000  
DEFINITION Homo sapiens BAC clone RP11-278G12 from 2, complete sequence.  
ACCESSION AC006369  
VERSION AC006369.3 GI:5757529  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 188359)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

```

MEDLINE
PUBMED
9847074
2 (bases 1 to 188359)
Threideh,J., Elliot,G. and Modde,T.
The sequence of Homo sapiens BAC clone RP11-278G12
Unpublished
3 (bases 1 to 188359)
Waterston,R.H.
Direct Submission
Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 188359)
Waterston,R.H.
Direct Submission
Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 188359)
Waterston,R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 21, 1999 this sequence version replaced gi:4337247.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0278G12
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-278G12;
actual end is at base position 188359 of RP11-278G12.
Location/Qualifiers
1. 188359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-278G12"
/clone_lib="RPCI-11"

FEATURES
source
1171..1580
/rpt_family="L1"
repeat_region
1583..1878
/rpt_family="Retroviral"
repeat_region
1879..2102
/rpt_family="L1"
repeat_region
2174..2728
/rpt_family="L1"
repeat_region
2901..3020
/rpt_family="MERL_type"
repeat_region
3049..3201
/rpt_family="MIR"
repeat_region
3249..3375
/rpt_family="MIR"
repeat_region
3415..3465
/rpt_family="Mariner"
repeat_region
4160..4333
/rpt_family="MaLR"
repeat_region
4334..4640
/rpt_family="Alu"
repeat_region
4641..4858
/rpt_family="MaLR"
repeat_region
4859..5075
/rpt_family="MIR"
repeat_region
5076..5386
/rpt_family="Alu"
repeat_region
5387..5428
/rpt_family="MIR"
repeat_region
5903..6215
/rpt_family="L2"
repeat_region
6242..6301
/rpt_family="L2"
repeat_region
6302..7290
/rpt_family="Retroviral"
repeat_region
7291..7678
/rpt_family="L2"
repeat_region
7695..7758
/rpt_family="(CCCA)n"
repeat_region
7759..8057
/rpt_family="MaLR"
repeat_region
9868..9934
/rpt_family="L2"
repeat_region
10653..10821
/rpt_family="MIR"
repeat_region
10923..11219
/rpt_family="Alu"
repeat_region
11553..11708
/rpt_family="MIR"
repeat_region
12552..12968
/rpt_family="L1"
repeat_region
12985..13283
/rpt_family="Alu"
repeat_region
13284..13306
/rpt_family="(TAAA)n"
repeat_region
13307..13325
/rpt_family="Alu"
repeat_region
14650..14717
/rpt_family="(CAT)n"
repeat_region
14724..14965
/rpt_family="MIR"
repeat_region
15679..16057
/rpt_family="MaLR"
repeat_region
16334..16552
/rpt_family="Alu"
repeat_region
17921..18131
/rpt_family="MaLR"
repeat_region
18132..18330
/rpt_family="L1"
repeat_region
18910..19251
/rpt_family="MER73-group"
repeat_region
19252..19397
/rpt_family="L1"
repeat_region
19398..19620

```





```
QY 1672 TTGAAGCTTTATGGGTTTGGGATTTTCTGTAGTTCTTATATATCCCTATATATATAT 1731
Db 112554 TTGAAGCTTTATGGGTTTGGGATTTTCTGTAGTTCTTATATATCCCTATATATATAT 112613
QY 1732 CTATATTGCAAAATTTTGACTGTCTAGCTACATGTTGGTAAGACACAGGCAAGTATTACT 1791
Db 112614 CTATATTGCAAAATTTTGACTGTCTAGCTACATGTTGGTAAGACACAGGCAAGTATTACT 112673
QY 1792 GTAAGCTTTATTTTAAAGTTAAATATATTTTACGTGCTTGGCTTTTATTTGCA 1851
Db 112674 GTAAGCTTTATTTTAAAGTTAAATATATTTTACGTGCTTGGCTTTTATTTGCA 112733
QY 1852 GAGTCTACATTTTATAGATTTTACATCAGATGTTGTCTTATTTCCATTTGGGATTTCCAT 1911
Db 112734 GAGTCTACATTTTATAGATTTTACATCAGATGTTGTCTTATTTCCATTTGGGATTTCCAT 112793
QY 1912 TGAAGCTGTGTATGCGTGTGTTTGGAAAAGTGTATTCATCTAGTTTCTTTCTTCA 1971
Db 112794 TGAAGCTGTGTATGCGTGTGTTTGGAAAAGTGTATTCATCTAGTTTCTTTCTTCA 112853
QY 1972 TCTGTTATCATCTTTTAAAGCAACCAATTAACGATTTGTAAGTGTAAAGGCACAGGTT 2031
Db 112854 TCTGTTATCATCTTTTAAAGCAACCAATTAACGATTTGTAAGTGTAAAGGCACAGGTT 112913
QY 2032 ACTCATGATGCTTCTGCGAGAGACTGTGGGCTACACCATATGTTTATTTGGAATATAGG 2091
Db 112914 ACTCATGATGCTTCTGCGAGAGACTGTGGGCTACACCATATGTTTATTTGGAATATAGG 112973
QY 2092 TATTTAGTACAGTACATCTTGCATTTACATAGGTACTTCAAGCAACCAATTAAGTA 2151
Db 112974 TATTTAGTACAGTACATCTTGCATTTACATAGGTACTTCAAGCAACCAATTAAGTA 113033
QY 2152 AATGATAAAGTGAAGTCTGTTGTTTATAGTAAACCAAGACCATTAAGATAAAGTATA 2211
Db 113034 AATGATAAAGTGAAGTCTGTTGTTTATAGTAAACCAAGACCATTAAGATAAAGTATA 113093
QY 2212 GCTAGAGAAATGCTTCTCTGAATGTACATGAGCCCTTAAGGTAAGAGATGATTTCCAT 2271
Db 113094 GCTAGAGAAATGCTTCTCTGAATGTACATGAGCCCTTAAGGTAAGAGATGATTTCCAT 113153
QY 2272 CTACTCTCATTTTGAATTAATCTCTTATGTTTGAAGGCTAGAACTGAGCCCTCTCTACT 2331
Db 113154 CTACTCTCATTTTGAATTAATCTCTTATGTTTGAAGGCTAGAACTGAGCCCTCTCTACT 113213
QY 2332 TTTGAAAAATGAACATGTGAGGTGAGTATTTTCTTTTAACTCAGCACTGATG 2391
Db 113214 TTTGAAAAATGAACATGTGAGGTGAGTATTTTCTTTTAACTCAGCACTGATG 113273
QY 2392 CCACCTCTCAGTGTGATTTCTGAGCATCTTCTGACTTGAACACCTTCTACAGCAAAAC 2451
Db 113274 CCACCTCTCAGTGTGATTTCTGAGCATCTTCTGACTTGAACACCTTCTACAGCAAAAC 113333
QY 2452 TCTTGCAAGTCCAGTTTCATCCCTGTAAAGCAAAATGCTTTTACGAGAAAGTGCCATA 2511
Db 113334 TCTTGCAAGTCCAGTTTCATCCCTGTAAAGCAAAATGCTTTTACGAGAAAGTGCCATA 113393
QY 2512 TAGACGAGATAAAGCAGCTANAAACGAGGCGAGTA 2546
Db 113394 TAGACGAGATAAAGCAGCTANAAACGAGGCGAGTA 113428
```

```
RESULT 2
HSM801810 2768 bp mRNA linear PRI 10-MAR-2001
LOCUS Homo sapiens mRNA; cDNA DKFZp434A0530 (from clone DKFZp434A0530);
DEFINITION complete cds.
ACCESSION AL136842
VERSION AL136842.1 GI:6807668
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS
1 (bases 1 to 2768)
Wiemann,S., Weil,B., Wellenreuther,R., Gassenhuber,J., Glassl,S.,
Ansong,W., Boecker,M., Bloeker,H., Bauersachs,S., Blum,H.,
Lauber,J., Duesterhoeft,A., Beyer,A., Koehrer,K., Strack,N.,
Mewes,H.W., Ottenwaelder,B., Obermaier,B., Tampe,J., Heubner,D.,
Wambutt,R., Korn,B., Klein,M. and Poustka,A.
Toward a Catalog of Human Genes and Proteins: Sequencing and
Analysis of 500 Novel Complete Protein Coding Human cDNAs
Genome Res. 11 (3), 422-435 (2001)
11230166
JOURNAL
PUBMED
REFERENCE
2 (bases 1 to 2768)
Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
Direct Submission
Submitted (15-JAN-2000) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp434A0530) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
source
1..2768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="178.4 cR from top of Chr2 linkage group"
/clone="DKFZp434A0530"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
969..1733
/gene="DKFZp434A0530"
969..1733
/gene="DKFZp434A0530"
/note="Cdc42 effector protein 3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66776.1"
/db_xref="GI:6807669"
/db_xref="SPTREMBL:Q9UKI2"
/translation="MPAKTPIYLKAAANNKKKFKLRDILSPDMISPPGLDFRHTIHI
GKEQHDVFDISFLQGNVYELLPNGQKALHQPFGHNEFFRANSTSDSVFTEPSPV
LKNAISLPTIGSQALMLPLLSPTVFNKQESFPAKLPLRSCPEVMEKAEKSSLL
ENGTVHQDTSWSSGSASQSSQGRDSSSSLSSEQYPDWPAEDMFDHTPTCELKRGKT
KSEESLSDLTGSLLSLQLDLGSLLDVLNVMDKNK"
2697..2702
polyA_signal
polyA_site 2711
BASE COUNT 705 a 717 c 676 g 670 t
ORIGIN
Query Match 85.1%; Score 2165.6; DB 9; Length 2768;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2183; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1 CGGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTCACGCCGCTCGGAGCGG 60
|||||
Db 557 CGGGGGCTAGCCCGGAGACCCGGCCACCGGCTGGGGCGCTTCACGCCGCTCGGAGCGG 616
QY 61 ATAATGCGGTGAGCAGGACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
|||||
Db 617 ATAATGCGGTGAGCAGGACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 676
QY 121 GATTGCGTGGCCCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
|||||
Db 677 GATTGCGTGGCCCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 736
QY 181 TGTCTCTGTTGTTGTTCAAGACTGGATGATGATGATGATGATGATGATGATGATG 240
|||||
```

Db 737 TGTCTCTGGTTGTGGTCAAGACTGGATGATGAATGAAGTGGCTCTCTAGGAAGCCTCAGTTGGC 796  
QY 241 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGCGAGCCACTTCTGAGGCCAC 300  
Db 797 CGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTTGCGAGCCACTTCTGAGGCCAC 856  
QY 301 TTGAGAAAAATGATGTGACAGTTCCCTATCAAAAAAGGATTCAGAAACATATACCATCTGTG 360  
Db 857 TTGAGAAAAATGATGTGACAGTTCCCTATCAAAAAAGGATTCAGAAACATATACCATCTGTG 916  
QY 361 AAGAAAGTGCCCTTCTCTCCCGCTTGCAAAATAGACATCTCTCAAAATCCAAATGCCAGC 420  
Db 917 AAGAAAGTGCCCTTCTCTCCCGCTTGCAAAATAGACATCTCTCAAAATCCAAATGCCAGC 976  
QY 421 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAGAAAGGAAAGAAATTTAAACTGAG 480  
Db 977 CAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAGAAAGGAAAGAAATTTAAACTGAG 1036  
QY 481 GGACATCTGTCTCTGATGATGATCAGTCCCGCTTGAGACTTTTCGCCACACCATCCA 540  
Db 1037 GGACATCTGTCTCTGATGATGATCAGTCCCGCTTGAGACTTTTCGCCACACCATCCA 1096  
QY 541 CATTTGGCAAGAGGGCCAGCAGATGTCTTTGGAGATATTTCCCTTCTTCAAGGGAACCTA 600  
Db 1097 CATTTGGCAAGAGGGCCAGCAGATGTCTTTGGAGATATTTCCCTTCTTCAAGGGAACCTA 1156  
QY 601 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTCCCTTCAAGGGAACCTA 660  
Db 1157 CGAGCTTTTACCTGGAACACAGGAGAAAGCACACCTGGGCCAGTCCCTTCAAGGGAACCTA 1216  
QY 661 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAAGCCTTCCCGCTGCT 720  
Db 1217 GTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTACAGAAAGCCTTCCCGCTGCT 1276  
QY 721 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATT 780  
Db 1277 CAAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTGCCCTTATT 1336  
QY 781 GTCACCAAGTACATTTAATTTCCAAACAGGAGTCCCTCGGCCAGCAAGCTGCGCCAGGCT 840  
Db 1337 GTCACCAAGTACATTTAATTTCCAAACAGGAGTCCCTCGGCCAGCAAGCTGCGCCAGGCT 1396  
QY 841 TAGCTGCGAGCCCGTCTATGGAGAAAGCTCAGGAGAAAGCAGTCTGTGGAGAAATGG 900  
Db 1397 TAGCTGCGAGCCCGTCTATGGAGAAAGCTCAGGAGAAAGCAGTCTGTGGAGAAATGG 1456  
QY 901 GACAGTCCACAGGAGACACCTCGTGGGGCTCAGCGGTTCTGCACTCAGTCCAGCCA 960  
Db 1457 GACAGTCCACAGGAGACACCTCGTGGGGCTCAGCGGTTCTGCACTCAGTCCAGCCA 1516  
QY 961 AGCAGAGACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGACTGGCCAGCCAGGA 1020  
Db 1517 AGCAGAGACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGACTGGCCAGCCAGGA 1576  
QY 1021 CATGTTTGACCATCCACCCCATGCGAGCTCATCAAGGGAAGACTAAGTCAGAGGAGTC 1080  
Db 1577 CATGTTTGACCATCCACCCCATGCGAGCTCATCAAGGGAAGACTAAGTCAGAGGAGTC 1636  
QY 1081 CCTCTCTGACCTTACAGGTTCCCTCCTCTCCCTGCAGCTTGATCTTGGGCCCTCAGTTT 1140  
Db 1637 CCTCTCTGACCTTACAGGTTCCCTCCTCTCCCTGCAGCTTGATCTTGGGCCCTCAGTTT 1696  
QY 1141 GGATGAGGTGCTGAATGTAATGGATAAAAAATAAGTAACAGATGCCAAGTCTTTTCTCCTTT 1200  
Db 1697 GGATGAGGTGCTGAATGTAATGGATAAAAAATAAGTAACAGATGCCAAGTCTTTTCTCCTTT 1756  
QY 1201 GGGGTAAAGGTACAAAAACAACTAACACAGTTGAGAGAAAGGCTTCCGGAGCTGTA 1260  
Db 1757 GGGGTAAAGGTACAAAAACAACTAACACAGTTGAGAGAAAGGCTTCCGGAGCTGTA 1816  
QY 1261 TTTGCAGTTTGTGTTGGGTTTCTAAAAATAATATCTTACAAAGTATTTTACCTGT 1320  
Db 1817 TTTGCAGTTTGTGTTGGGTTTCTAAAAATAATATCTTACAAAGTATTTTACCTGT 1876

QY 1321 TATGCCCTGTGTTGCAAAAAACRAATTTAGAAAAAACAAACAAAGCAAAACCTATCTTGGCAA 1380  
Db 1877 TATGCCCTGTGTTGCAAAAAACRAATTTAGAAAAAACAAACAAAGCAAAACCTATCTTGGCAA 1936  
QY 1381 AAAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCAATGGTGATGTTGCGGCTCACATAT 1440  
Db 1937 AAAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCAATGGTGATGTTGCGGCTCACATAT 1996  
QY 1441 TGTTTGCAGACACACAGAAATCTGGCTTGGCCAGGATTTGGCACTAGCTATGAAGGCTG 1500  
Db 1997 TGTTTGCAGACACACAGAAATCTGGCTTGGCCAGGATTTGGCACTAGCTATGAAGGCTG 2056  
QY 1501 AGCGAGTCACATTAAGGAACCTTCACGGAACCTTTATAGCACTCCGACATTTTCTGAGCAAG 1560  
Db 2057 AGCGAGTCACATTAAGGAACCTTCACGGAACCTTTATAGCACTCCGACATTTTCTGAGCAAG 2116  
QY 1561 AGGAAGTCAAAATTTATTTAACACCTTAAGCCTTTTGTAGACTCTTTTCTATATATATGCT 1620  
Db 2117 AGGAAGTCAAAATTTATTTAACACCTTAAGCCTTTTGTAGACTCTTTTCTATATATATGCT 2176  
QY 1621 TAGGCTCACCATAGCGAATCTCCAGTGTAAAAACCTTTTCTGTTTTCACATTTGAACCTTT 1680  
Db 2177 TAGGCTCACCATAGCGAATCTCCAGTGTAAAAACCTTTTCTGTTTTCACATTTGAACCTTT 2236  
QY 1681 ATGGGTTTTGGGGATTTTCTTGTAGTTCTTATATATATATATATATATATATATATGCT 1740  
Db 2237 ATGGGTTTTGGGGATTTTCTTGTAGTTCTTATATATATATATATATATATATATATGCT 2296  
QY 1741 AAAATTTGACTGTGAGTACATGTTGGTAAGACACAGGCAAGATTTACTGTAACTAAG 1800  
Db 2297 AAAATTTGACTGTGAGTACATGTTGGTAAGACACAGGCAAGATTTACTGTAACTAAG 2356  
QY 1801 TTATTTTAAAGTTAAAAATATATTTTACGTGCTTTGGCTTTTATTCAGAGTCTACA 1860  
Db 2357 TTATTTTAAAGTTAAAAATATATTTTACGTGCTTTGGCTTTTATTCAGAGTCTACA 2416  
QY 1861 TTTATAGATTTCTACATCAGATGTTGTCACTTATTTCCATTTGGGATTCATTTGAAGCTG 1920  
Db 2417 TTTATAGATTTCTACATCAGATGTTGTCACTTATTTCCATTTGGGATTCATTTGAAGCTG 2476  
QY 1921 TGTATGTCGTGTTTGGAAAAAGTGTATTCATCTAGTTTCTTCTCATCTGTTATC 1980  
Db 2477 TGTATGTCGTGTTTGGAAAAAGTGTATTCATCTAGTTTCTTCTCATCTGTTATC 2536  
QY 1981 ATACTTTTAAACAGCAACCAATAACGGATTTGTAAGTGTAAAGCAAGGTTACTCATGAT 2040  
Db 2537 ATACTTTTAAACAGCAACCAATAACGGATTTGTAAGTGTAAAGCAAGGTTACTCATGAT 2596  
QY 2041 GCTTCTGCAGAGACTGTGGGCTACACACATATGTTTATTTGGAAATATAGGATTTTAGT 2100  
Db 2597 GCTTCTGCAGAGACTGTGGGCTACACACATATGTTTATTTGGAAATATAGGATTTTAGT 2656  
QY 2101 ACAGTACATCTGCTTACATAGTACTTCAAGCAACACATAAAGTAATGATAAA 2160  
Db 2657 ACAGTACATCTGCTTACATAGTACTTCAAGCAACACATAAAGTAATGATAAA 2716  
QY 2161 GTGAAGTGTGCTTGTATAGTAAATAAACAGCAAGCAAGCAAGTAAGTATAG 2212  
Db 2717 AAAG 2768

RESULT 3  
BC019270  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BC019270 1892 bp mRNA linear PRI 19-DEC-2001  
Homo sapiens, clone MGC:3481 IMAGE:2823261, mRNA, complete cds.  
BC019270  
BC019270.1 GI:17939542  
MGC.  
Homo sapiens.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.







QY	1381	AAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCATGGTGATGTTCCGGCTCACATAT	1440
DB	1406	AAAAAGGAAGTGAGTCAGAGCCCATTTTCAGGAGGCATGGTGATGTTCCGGCTCACATAT	1465
QY	1441	TGTTTCAGACACACAAGAAATCTGGCTTGGCCAGGATTGGCACTAGCTATGAAGGGCTG	1500
DB	1466	TGTTTCAGACACACAAGAAATCTGGCTTGGCCAGGATTGGCACTAGCTATGAAGGGCTG	1525
QY	1501	ACCGAGTCACATTAAGGAACCTTCACGGAACCTTATAGCACICCGACATTTCTGAGCAAG	1560
DB	1526	ACCGAGTCACATTAAGGAACCTTCACGGAACCTTATAGCACICCGACATTTCTGAGCAAG	1585
QY	1561	AGGAAGTCAAAAATTATTTAACACCTTAAGCCCTTTTGTAGACTCTTTTCTATATATTGCT	1620
DB	1586	AGGAAGTCAAAAATTATTTAACACCTTAAGCCCTTTTGTAGACTCTTTTCTATATATTGCT	1645
QY	1621	TAGGCTCACCATAGCGAATTCTCCAGTGTAAACCTTTTCTGTTTTTCACATTTGAACCTTT	1680
DB	1646	TAGGCTCACCATAGCGAATTCTCCAGTGTAAACCTTTTCTGTTTTTCACATTTGAACCTTT	1705
QY	1681	ATGGGTTTTGGGGATTTTCTTGTAGTTCCTATATATATATATATATATATATATATATATGC	1740
DB	1706	ATGGGTTTTGGGGATTTTCTTGTAGTTCCTATATATATATATATATATATATATATATATGC	1765
QY	1741	AAAAATTTGACTGTGAGCTACATGTTGGTAAGACACAGGCAAGTATTACTGTAACCTAAG	1800
DB	1766	AAAAATTTGACTGTGAGCTACATGTTGGTAAGACACAGGCAAGTATTACTGTAACCTAAG	1825
QY	1801	TTATTTTTAAAGTTAAAAATATATTTTACGTGCCTTTTGGC	1840
DB	1826	TTATTTTTAAAGTTAAAAATATATTTTACGTGCCTTTTGGC	1865

RESULT 4	
AC022409	
LOCUS	93663 bp DNA linear PRI 23-AUG-2001
DEFINITION	Homo sapiens chromosome 19 clone CTB-175p5, complete sequence.
ACCESSION	AC022409
VERSION	AC022409.7 GI:15281220
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 93663) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Unpublished 2 (bases 1 to 93663) DOE Joint Genome Institute. Direct Submission Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA 3 (bases 1 to 93663) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Submitted (08-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA 4 (bases 1 to 93663) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Aug 23, 2001 this sequence version replaced gi:11120761. Draft sequence produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www-shgc.stanford.edu Quality: Phrap Quality >=40 99.4% of Sequence; Estimated Total Number of Errors is 0.2. STS Content: SHGC-150482 G60581. Location/Qualifiers
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	







QY 1073 GAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCCTCCTGAGCTTGATCTGGGCC 1132  
Db 721 GAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCCTCCTGAGCTTGATCTGGGCC 780  
QY 1133 TCACCTTTTGGATGAGGTGCTGAATGTAATGATGATAAATAAAGTAACAAGATGCCAATTT 1192  
Db 781 TCACCTTTTGGATGAGGTGCTGAATGTAATGTAATGATAAATAAAGTAACAAGATGCCAATTT 840  
QY 1193 TTTCCTTTGGGTAAGGTACAAAACAACTAACCCACAGTTGAAGAGAAGGGCTTCCG 1252  
Db 841 TTTCCTTTGGGTAAGGTACAAAACAACTAACCCACAGTTGAAGAGAAGGGCTTCCG 900  
QY 1253 GAGCTGTA-TTTGCACTTTTGTGTTGGGTTTCTAAATAA 1292  
Db 901 GAGCTGATTTTGCATTTTGTGTTGGGTTTCTAAATAA 941

RESULT 7  
AC091332/c  
LOCUS  
DEFINITION Mus musculus clone RP23-12K1, WORKING DRAFT SEQUENCE, 9 unordered  
pieces.  
AC091332  
VERSION AC091332.3 GI:22325102  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 216905)  
Birren,B., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-12K1  
Unpublished  
2 (bases 1 to 216905)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Tesfaye,S.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Submitted (17-APR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 216905)  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,

Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 20, 2002 this sequence version replaced gi:14209739.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L13294  
Center clone name: 12\_K1  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 214909 bases at least Q40  
Consensus quality: 215495 bases at least Q30  
Consensus quality: 215817 bases at least Q20  
Insert size: 210000; agarose-fp  
Insert size: 216105; sum-of-contents  
Quality coverage: 13.7 in Q20 bases; agarose-fp  
Quality coverage: 13.3 in Q20 bases; sum-of-contents

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 78339: contig of 78339 bp in length  
\* 78340 78439: gap of 100 bp  
\* 78440 79696: contig of 1257 bp in length  
\* 79697 79796: gap of 100 bp  
\* 79797 83334: contig of 3538 bp in length  
\* 83335 83434: gap of 100 bp  
\* 83435 89559: contig of 6125 bp in length  
\* 89560 89659: gap of 100 bp  
\* 89660 97693: contig of 8034 bp in length  
\* 97694 97793: gap of 100 bp  
\* 97794 113828: contig of 16035 bp in length  
\* 113829 139314: contig of 25386 bp in length  
\* 139315 139414: gap of 100 bp  
\* 139415 190921: contig of 51507 bp in length  
\* 190922 191021: gap of 100 bp  
\* 191022 216905: contig of 25884 bp in length.  
----- Location/Qualifiers  
1. 216905  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-12K1"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. 78339  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
78440. 79696  
/note="assembly\_fragment"  
79797. 83334  
/note="assembly\_fragment"  
83435. 89559  
/note="assembly\_fragment"  
89660. 97693

FEATURES  
Source  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature





```
QY 2033 CTCATGATGCTTCTGCAGAGACTGTGGGCTACACCACATATGTTATTGGAATATAGGT 2092
Db 203653 -----TGATGCTTCTGTGGAGACTG-----GAGTGAGGTAGGT 203620
QY 2093 ATTTAGTACAGTACATACACTTGCATTACATAGTACTTCAAGCAACACATAAAAAAGTAA 2152
Db 203619 ATTTAAGGACAGCTCATGTTCCATTACTTGGGTACATCATACCACCGCAATAAAGAGTAA 203560
QY 2153 ATGATAAAGTGAACCTTGCTGTTTATAGTAAATAAACAAGA--CCATAAGAGAGAATAAGTAT 2210
Db 203559 ATGCTAAAATGAACGTTGTCATTATAGTCAATAAACAAGAACCCGAAAAGATTTTAAGACC 203500
QY 2211 AGCTAGAGAAATTCCTCTCTGAAATGTACATGAGCCCTTAAGGTAAGAGATGATTCCCA 2270
Db 203499 ATCTAGTAAAGTCCCTAAAATGTACCCACACCTTTAACATAAGCCCAAGAGATGACTGCAG 203440
QY 2271 TCTACTCTCATTT 2283
Db 203439 TCTTGACTTTT 203427

RESULT 8
AC055705
LOCUS
DEFINITION Mus musculus clone RP23-17H14 strain C57BL6/J, *** SEQUENCING IN
PROGRESS ***, 91 unordered pieces.
AC055705
VERSION AC055705.11 GI:21306615
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 292420)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 292420)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (18-APR-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
3 (bases 1 to 292420)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (01-JUN-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA
On Jun 1, 2002 this sequence version replaced gi:17976429.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcgg.org/Sequence/mouse.html
Contact: hpcgcmendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ACF
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 274929 at least Q20
*Consensus quality: 267475 at least Q30
*Consensus quality: 253492 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 290620 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 3.7 x in Q20 bases; sum-of-contigs estimation
-----
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 91 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 29525: contig of 29525 bp in length  
\* 29526 29545: gap of unknown length  
\* 29546 49030: contig of 19485 bp in length  
\* 49031 49050: gap of unknown length  
\* 49051 64774: contig of 15724 bp in length  
\* 64775 64794: gap of unknown length  
\* 64795 76739: contig of 11945 bp in length  
\* 76740 76759: gap of unknown length  
\* 76760 87926: contig of 11167 bp in length  
\* 87927 87946: gap of unknown length  
\* 87947 99465: contig of 11519 bp in length  
\* 99466 99485: gap of unknown length  
\* 99486 110785: contig of 11300 bp in length  
\* 110786 110805: gap of unknown length  
\* 110806 122033: contig of 11228 bp in length  
\* 122034 122053: gap of unknown length  
\* 122054 133409: contig of 11356 bp in length  
\* 133410 133429: gap of unknown length  
\* 133430 140278: contig of 6849 bp in length  
\* 140279 140298: gap of unknown length  
\* 140299 148523: contig of 8225 bp in length  
\* 148524 148543: gap of unknown length  
\* 148544 156296: contig of 7753 bp in length  
\* 156297 156316: gap of unknown length  
\* 156317 163185: contig of 5869 bp in length  
\* 163186 163205: gap of unknown length  
\* 163206 169646: contig of 6441 bp in length  
\* 169647 169666: gap of unknown length  
\* 169667 174627: contig of 4961 bp in length  
\* 174628 174647: gap of unknown length  
\* 174648 180224: contig of 5577 bp in length  
\* 180225 180244: gap of unknown length  
\* 180245 186004: contig of 5760 bp in length  
\* 186005 186024: gap of unknown length  
\* 186025 190927: contig of 4903 bp in length  
\* 190928 190947: gap of unknown length  
\* 190948 195412: contig of 4465 bp in length  
\* 195413 195432: gap of unknown length  
\* 195433 199927: contig of 4495 bp in length  
\* 199928 199947: gap of unknown length  
\* 199948 204193: contig of 4246 bp in length  
\* 204194 204213: gap of unknown length  
\* 204214 208850: contig of 4637 bp in length  
\* 208851 208870: gap of unknown length  
\* 208871 210743: contig of 1873 bp in length  
\* 210744 210763: gap of unknown length  
\* 210764 213706: contig of 2943 bp in length  
\* 213707 213726: gap of unknown length  
\* 213727 216378: contig of 2652 bp in length  
\* 216379 216398: gap of unknown length  
\* 216399 219944: contig of 3546 bp in length  
\* 219945 219964: gap of unknown length  
\* 219965 221643: contig of 1679 bp in length  
\* 221644 221663: gap of unknown length  
\* 221664 223040: contig of 1377 bp in length  
\* 223041 223060: gap of unknown length  
\* 223061 224619: contig of 1559 bp in length  
\* 224620 224639: gap of unknown length  
\* 224640 226625: contig of 1986 bp in length  
\* 226626 226645: gap of unknown length  
\* 226646 226730: contig of 85 bp in length  
\* 226731 226750: gap of unknown length  
\* 226751 228130: contig of 1380 bp in length  
\* 228131 228150: gap of unknown length  
\* 228151 229444: contig of 1294 bp in length





QY 1132 CTCACCTTTTGGATGAGGTGCTGAATGTAATGGATAAAAAAATAAGTAACAAGATGCCAACTT 1191  
Db 75562 TTTACCTTTTGGATGAAGTCTCTGAATGTCATGGATAAAAAAATAAGTAGCCAGGATCAGCTC 75621  
QY 1192 TTTTCCTTTTGGGTAAAGGTAC-----AAAAACAACCTAACCCACA 1232  
Db 75622 TTTTCTTGGGAGTAAGAGGTACCAAAATAAGCAGAAAAACAATAAATCACTAACCCACA 75681  
QY 1233 GTTGAAGAGAGAGGCTTCCGGAGCTG----TATTTGACGTTTGTGTGGTGTCTTCTAAA 1288  
Db 75682 GCTGATGGAAGAGAGCTGCTCTGACTGTTTATTTGACGACATGTGATGCTATTTCTAAA 75741  
QY 1289 ATAA-----TATTTCTTACAAAGTATTTTACCTGTATGCCCCCTGTTTGCAAAAAC-- 1340  
Db 75742 ATAAAGCTCTGTTCTTACAAATATTTTAAACCATCATGCCCTGTTTGCAAAAATTT 75801  
QY 1341 -----AATTTAGAAAAAACAACAAGCAAAACCTATCTTGGCAAAAAAAGGAAG 1390  
Db 75802 TGAACAACAATAATTTTAAATAATAAAAAAACAACAACAACAACCTGTTCTGG 75861  
QY 1391 TGAGTCAGAGCCCATTTTTCAGGAGGCATTTGTTGATGTTTGGCTCACATATGTTTG--CA 1448  
Db 75862 CAAAAAAGAAAAAAGAAAAAGAAAGGAGGAGTGTGAGTCAAGCTCATTTTCAACAG 75921  
QY 1449 GACACACAAGAAATCTGGCTTGGCCAGGATTTGCACTAGCTATGAAGGGCTGAGCGAGTC 1508  
Db 75922 GCATTTGCTGATGCTCATACATGCTCAGTTTGTGTAGACATGCAGAAATCCAGGCTGT 75981  
QY 1509 ACATTAAGGAACCTTACGGAACCTTTATAGCACTCCGACATTTTCTGAGCAAGGAAGTC 1568  
Db 75982 CTAAGAGGCTGGGTCCATGGTCTCTCCAGGCAATTCATTGATGTTATGAGAGGAA 76041  
QY 1569 A-AAATTTATTTAACACCTAAGCCTTTTGTAGACTCTTTTCT-----ATATATTGCTT 1621  
Db 76042 ACTTACTTATTTGGACCTACGCTTTTTCCTAGCCTTTTGTATATTTGTATCTTTT 76101  
QY 1622 AGGCTCACCATAGCAATTTCTCCAGTGTAAACCTTTTCTGTTTTCACATTTG---AACT 1678  
Db 76102 GGGTTCACCACAGCGAAGTCTCCAGGCTATGATTTTACTCTTTCATATTCGAGCAACC 76161  
QY 1679 TTATGGGTTTTTGGGATTTTCTT-----GTAGTTCTTATATATATATATATATATCT 1733  
Db 76162 TTACAGGATTTGGGATTTTTCCTTGTGGTTCCTGATATATATATATATATATATCT 76221  
QY 1734 ATATTGCAAAATTTTACTGTCAGCTACATGTTGGTAAGACACAGCAAGGATTTACTGT 1793  
Db 76222 ATGTTGCAAAAGTTTACTGTCAGCTACATGTTGGTAAACAAAGCAAGGATTTACTGT 76281  
QY 1794 AACTAAGTTATTTTAAAGTTAAATATATTTTACGTCCTTTTGGCTTTTATTTATTCAGA 1853  
Db 76282 AACTAAGTTATTTTAAAGTTGAATATATTTTATGTCCTTTTGGCTTTTATTTGCAA 76341  
QY 1854 GTCTACATTTTATAGATTCTACATCAGATGT--TGTCACCTTATTTCCATTTGGGATTCAT 1911  
Db 76342 GTCTACATTTTATACATAATACATCAAGTGTATTTGTCGTGTTTCCATTCGGGCTTCCT 76401  
QY 1912 TGAAGCTGTGTATGTCGTGTTTGGAAAAGTGTATTTACATCTTACTTTTCTTCTCA 1971  
Db 76402 TGCAAGCTACGCTGTGTGTTTGAACAACAACCAACCAACCAACCTATCGGTTATCAGCCT 76461  
QY 1972 TCTGTATATACATCTTTTAAACAGCAACCAATAACGGATTTGTAAGTGTAAAGGCACAGTT 2031  
Db 76462 GACCTGTTGCTCGGGTGTACCTCTCCAGTAGGAGGCACAGG-----76507  
QY 2032 ACTCATGATGCTTCTGACAGAGACTGTGGGCTACACCATATGTTATTTGGAAATATAGG 2091  
Db 76508 -----TGATGCTTCTGTGGAGACTG-----GAGGTGAGGTAGG 76540  
QY 2092 TATTTTAGTACAGTACATCTTGCATTTACATAGGTACTTCAAGCAACACATAAAAACTA 2151  
Db 76541 TATTTAAGGACAGCTCATGTTTACATTTACTTGGGTACATCATACCAACGCAATAAAGAGTA 76600  
QY 2152 AATGATAAAGTGAACCTGCTTGTGTTTATAGTAATAAACAAGA--CCATAAGAGAATAAGTA 2209

Db 76601 AATGCTAAATGAACGCTGTTTCATTTATAGTCAATAAACAAGAACCGAAAGATTTTAAGAC 76660  
QY 2210 TAGCTAGAGAAATTTGCTTCTCTGAAATGTACATGAGCCCTTAAGCTAAGAGATGATTTCC 2269  
Db 76661 CATCTAGTAAAGTCTCTAAATGTACCCACACCTTTAACATAAGCAAGAGATGACTGCA 76720  
QY 2270 ATCTACTCTCATTT 2283  
Db 76721 GTCTTGACTTTTTT 76734

RESULT 9  
BC034714  
LOCUS  
DEFINITION  
Mus musculus, RIKEN cDNA 320001F04 gene, clone MGC:11796  
IMAGE:3595315, mRNA, complete cds.  
ACCESSION  
BC034714  
VERSION  
BC034714.1 GI:21961177  
KEYWORDS  
MGC  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1481)  
Strausberg, R.  
Direct Submission  
Submitted (24-JUL-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 16 Row: j Column: 15.

FEATURES  
source

Location/Qualifiers  
1..1481  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:11796 IMAGE:3595315"  
/tissue\_type="Mammary tumor. C3(1)-Tag model. Infiltrating  
ductal carcinoma. 5 month old virgin mouse."  
/clone\_lib="NCI\_CGAP\_Mam6"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
456..1220  
/codon\_start=1  
/product="RIKEN cDNA 320001F04 gene"  
/protein\_id="AAH34714.1"  
/db\_xref="GI:21961178"  
/db\_xref="LocusID:68025"  
/translation="MPAKTPIYLKAAANKKKFKLRDILSPDMISPPPLGDFRHTIHI  
GKEGQHDVFGDISFLOGNYELLPGNQEKASHGQFPNGHNDFFRANSTDSMFETPSV  
LKNALSLPTIGGSOALMLPLLSPTVFHKSQESFGRPKLRLSCBPVMEKVOEQSSL  
ENGAVHQGDTSWGSSGSGSQSSQGRDSSHSSLSSEQSDWPADDMFEHPASCELVKST  
KSEESFDLTGSLSLQLDLGSLDLVNVMDKNK"

CDS

BASE COUNT 405 a 394 c 353 g 329 t  
ORIGIN



Query Match		30.2%;	Score 769.8;	DB 10;	Length 1481;
Best Local Similarity		75.2%;	Pred. No. 7.6e-177;		
Matches 1062;		Conservative 0;	Mismatches 297;	Indels 54;	Gaps 6;
QY	20	CGGCCACCGGCTGGGGCGCCTT	CACGGCGTCTCGGAGCGGATAT	CGCGGTGAGCAGGCA	79
Db	60	CGGCTGCCTCCCGGAGACGCCG	CGCGAGTCCCGGAGCAGATTCT	CGCGGTGAACCGTTG	119
QY	80	CCAGCGCGGACAGACTCGGCT	GGATCTGGC-----CACAGCGG	CAGGATTCGCT	128
Db	120	CGGCCCTCTGTTCTGCGCAC	CGCAGGAGTGTTCGCGCTCGG	AGCGACCCGG	179
QY	129	CGCCCGCGGAGGCGCGGCG	CAGCGGCTCGTCAAGCGGCGG	CGGCTTGTCTCG	188
Db	180	GCCGACGCCGCGCTCCGGC	CAGAGCTAGGAGCCTTAGAG	ACTGCTGTTCTCCGAG	239
QY	189	TTGTGGTCAAGACTGGATGAT	GTAACTGGCTCTTAGGAAGCCT	CACTTGGCCGTAACT	248
Db	240	CTGTGGAAGGACTCGATAAT	GTAACTGGCTCTCCGGAAGAC	CGCGGTGGCCCTGCGCT	299
QY	249	CAGGAAGGTTCTCTTTGAC	CCCCCATCTCATTTGGAAGCC	CACTTGAAGCCACTTGAG	308
Db	300	GAGCAAGATCTCCAGACCC	TGTCTCGTTTGTAGAGCC	CACTTGAGAAG-----A	348
QY	309	AATGATGTGACAGTTCTT	ATCAAAAAGGA--TTCAGAA	ACATATACCATCTGTGA	365
Db	349	AATGATGTGACAAATCTT	ATGAAAAGGACGTTTGG	AAGTGTAGGATACCCGT	408
QY	366	AGTGGCCCTTCTCCCGCT	TGCAAAATAGACATTTCT	CAAAATCCAAAATGCC	425
Db	409	GGTCACTCTTTCCCTGGT	CTGTGAATAGACATTTCT	CAAAATCCAAAATGCC	468
QY	426	CCCCAATTTACCTGAAAG	CAGCCCAATTAACAAGAA	AGGAAAGAAATTTAA	485
Db	469	CCCCAATTTACCTGAAAG	CAGCTAACAAAGAAAGG	AAAGAAATTTAAACT	528
QY	486	TTCTGTCTCCTGATATG	ATCAGTCCCGCTTGAGAC	TTTCGACACCATCC	545
Db	529	TTCTGTCCCTGATATG	ATCAGTCCCTCGGGGAC	TTTCGTCACACTAT	588
QY	546	GCAAAGAGGCGCAGCAG	TGCTTTTGGAGATATTT	CCCTTTCTCAAGGG	605
Db	589	GCAAAGAGGCGCAGCAG	TGCTTTTGGGATATTT	CCCTTTCTCAAGGG	648
QY	606	TTTTACCTGGAAACCA	GAGAGAAAGCACACCT	GGGCCAGTTCCCTG	665
Db	649	TCTTGCCGGGAACCA	GAGAGAGGCACACTG	TGGCCAGTTCCCGG	708
QY	666	TCCGGGCCAACACAG	CACCTCGGACTCTGTG	TTCACAGAAACGCG	725
Db	709	TCCGGGCCAACACAG	CACCTCGGACTCAATG	TTTACAGAAACAC	768
QY	726	ATGCCATCTCCCTCC	CGACCATTTGAGGATC	CCCAAGCTCTATGTT	785
Db	769	ATGCCATCTCCCTCC	CGACCATTTGAGGATC	CCCAAGCTCTATGTT	828
QY	786	CAGTGACATTTAAT	TCCAAACAGGAGTCT	TTCGGGCCAGCAAG	845
Db	829	CAGTGACATTTAAT	TCCAAACAGGAGTCT	TTCGGGCCAGCAAG	888
QY	846	GCGAGCCCGTCA	TGGAGGAAAAGCTC	AGGAGAAAGCAGTCT	905
Db	889	GTGAGCCTGTCAT	GGAAGAGAAAGTTC	AGGAGCAAGTAGTCT	948
QY	906	TCCACAGGAGAC	ACCTCGTGGGCTCC	AGCGGTTCTGATCT	965
Db	949	TCCACAGGAGAC	ACCTCGTGGGCTCC	AGCGGTTCTGATCT	1008
QY	966	GAGACAGCCTCT	CCAGCCTGTCCGAAC	AGTACCCCGACTGG	1025
Db	1009	GCGACAGCCTCT	CCAGCCTGTCCGAAC	AGTACCCCGACTGG	1068

QY	1026	TTGACCATCCACCCCATGCGAGCTCATCAAGGGAAGACTAAGTCAGAGGAGTCCCTCT	1085
Db	1069	TTGAGCATCTGCTGCTCTGAGCTTGTGAAGTCAAGACTAAATCAGAAGAGTCTCTCT	1128
QY	1086	CTGACCTTACAGGTTCCCTCTCTCCCTGCAGCTTGATCTTGGGCCCTCACTTTTGGATG	1145
Db	1129	CTGACCTGACCGGCTCCCTGCTCTCCCTGCAGTTGGATCTTGGGCCCTCACTTTTGGATG	1188
QY	1146	AGTGCTGAATGTAATGGATAAAATAAGTAACAAGATGCCAACTTTTCTCTTGGGGT	1205
Db	1189	AAGTCTGAATGTATGGATAAAATAAGTAGCCAGGAGTCAGCTCTTTTCTTGGGAGT	1248
QY	1206	AAAAGGTAC-----AAAAACAACTAACACAGTTGAAGAGAAGG	1246
Db	1249	AAGAGTACCAAAATAAGCAGAAAAACAAAAATCAACTAACCCACAGCTGATGGAAGAG	1308
QY	1247	CTTCCGGAGCTG---TATTTGCAGTTTGTGTGGGTTTCTAAATAA-----TATT	1296
Db	1309	CTGCTCTGACTGTTTATTTATTTGCAGCCATGTGATGATTTCTAAATAACGCTCCTGTT	1368
QY	1297	CTTACAAAAGTATTTTACCTGTTATGCCCTGTTGCAAAACAATTTAGAAAAACA	1356
Db	1369	CCTACAAAATATTTTAAACCATCATGCCCTGTTGCAAAATTTGAAAAACAATAATT	1428
QY	1357	ACAAAGCAAAACCTATCTTGGCAAAAAGGAA	1389
Db	1429	TTTAAAAAATAAAAAAAAAAAAAAAAAAAAA	1461

RESULT 10  
AF094521

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

AF094521 Homo sapiens MSE55-related protein (UB1) mRNA, complete cds.  
AF094521.1 GI:3834632

765 bp linear PRI 06-NOV-1998

Homo sapiens.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 765)  
Alberts, A.S., Bouquin, N., Johnston, L.H. and Treisman, R.

Analysis of RhoA-binding proteins reveals an interaction domain conserved in heterotrimeric G protein beta subunits and the yeast response regulator protein Skn7

J. Biol. Chem. 273 (15), 8616-8622 (1998)

2 (bases 1 to 765)  
Alberts, A.S.

MSE55-related Cdc42-binding protein  
Unpublished

3 (bases 1 to 765)  
Alberts, A.S.

Direct Submission  
Submitted (24-SEP-1998) Cancer Center, University of California, San Francisco, 2340 Sutter St., San Francisco, CA 94115, USA

Location/Qualifiers  
1. .765

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/tissue\_type="brain"  
/dev\_stage="fetal"

1. .765  
/gene="UB1"

1. .765  
/gene="UB1"

/note="Cdc42 binding protein; similar to Homo sapiens serum protein MSE55; SwissProt Accession Number Q00587 and unknown protein encoded by the sequence presented in GenBank Accession Number AF001436; contains CRIB-motif"

/codon\_start=1  
/product="MSE55-related protein"

```
/protein_id="AAC71773.1"
/db_xref="GI:3834633"
/translation="MPAKTYIYLKAANNKKGKFKLRDILSPDMISPLGLDFRHTIHI
GKEGHDVFGDISFLOGNYELLPGNOEKALHQFGPHNEFFRANSTSDSVFTETPSPV
LKNALISLPTIGGSQALMLPLSPVTFNSKQESFGPAKLRLSCEPVMEKAEKSSLL
ENGIVHQGDTSWGSSGSASQSSQGRDSSHSSLSSEQYPDWPAEDMFDHPTFCELIKGT
KSESLSDLTGSLFLQLDLGSLLLDEVLVNMDRKN"
BASE COUNT      203 a      221 c      178 g      163 t
ORIGIN
Query Match      29.9%; Score 761.8; DB 9; Length 765;
Best Local Similarity 99.7%; Pred. No. 6.1e-175;
Matches 763; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 413 ATGCCAGCCCAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTT 472
Db 1 ATGCCAGCCCAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGGAAAGAAATTT 60
QY 473 AAACGTGAGGACATCTGTCTCCTGATATGATCAGTCCCGCGTTGGAGACTTTTCGCCAC 532
Db 61 AAACGTGAGGACATCTGTCTCCTGATATGATCAGTCCCGCGTTGGAGACTTTTCGCCAC 120
QY 533 ACCATCCACATTTGGCAAAGAGGGCCAGCACGATGTCTTTGGAGATATTTCCCTTTCTCAA 592
Db 121 ACCATCCACATTTGGCAAAGAGGGCCAGCACGATGTCTTTGGAGATATTTCCCTTTCTCAA 180
QY 593 GGAACACTACGAGCTTTTACCTGGAACACGAGGAGAAAGCACACTGGGCCAGTTCCCTGGG 652
Db 181 GGAACACTACGAGCTTTTACCTGGAACACGAGGAGAAAGCACACTGGGCCAGTTCCCTGGG 240
QY 653 CATAATGAGTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTCACAGAAACGCCCTCC 712
Db 241 CATAATGAGTTCTTCCGGGCCAACAGCACCTCGGACTCTGTGTTCACAGAAACGCCCTCC 300
QY 713 CCGGTGCTCAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTG 772
Db 301 CCGGTGCTCAAAATGCCATCTCCCTCCCGACCATTTGGAGGATCCCAAGCTCTCATGTTG 360
QY 773 CCCTTATTGTACCAAGTGACATTTTAAATCCAAACAGGAGTCTTCCGGGCCAGCAAGCTG 832
Db 361 CCCTTATTGTACCAAGTGACATTTTAAATCCAAACAGGAGTCTTCCGGGCCAGCAAGCTG 420
QY 833 CCCAGGCTTAGCTCGGAGCCCGTCTATGGAGGAAAGACTCAGGAGAAAGCAGTCTGTTG 892
Db 421 CCCAGGCTTAGCTCGGAGCCCGTCTATGGAGGAAAGACTCAGGAGAAAGCAGTCTGTTG 480
QY 893 GAGAATGGACAGTCCACCAGGGAGACACCTCGTGGGGCTCCAGCGGTTCTGCATCTCAG 952
Db 481 GAGAATGGACAGTCCACCAGGGAGACACCTCGTGGGGCTCCAGCGGTTCTGCATCTCAG 540
QY 953 TCCAGCCAAAGGAGAGACAGCCACTCTCCAGCCTGTCCGACAGTACCCCGACTGGCCA 1012
Db 541 TCCAGCCAAAGGAGAGACAGCCACTCTCCAGCCTGTCCGACAGTACCCCGACTGGCCA 600
QY 1013 GCCGAGGACATGTTTGACCATCCCAACCCCATGCGAGCTCATCAAGGAAAGACTAAGTCA 1072
Db 601 GCCGAGGACATGTTTGACCATCCCAACCCCATGCGAGCTCATCAAGGAAAGACTAAGTCA 660
QY 1073 GAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCTCCCTGACAGTTGATCTTGGGCC 1132
Db 661 GAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCTCTCTGACAGTTGATCTTGGGCC 720
QY 1133 TCACCTTTTGGATGAGGTGCTGAATGTAATGGATAAAATAAGTAA 1177
Db 721 TCACCTTTTGGATGAGGTGCTGAATGTAATGGATAAAATAAGTAA 765
```

```
RESULT 11
AC090491/c
LOCUS      AC090491      204019 bp      DNA      linear      HTG 24-FEB-2001
DEFINITION Mus musculus clone RP23-12K1, *** SEQUENCING IN PROGRESS ***, 43
unordered pieces.
ACCESSION AC090491
```

```
AC090491.1 GI:13122740
HTG; HTGS_PHASE1.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 204019)
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Balijs,V.,
Dedhia,N.N., de la Bastide,M., Huang,E.N., King,L., Kirchoff,K.A.,
Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Vil,M.D.
and Zutavern,T.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 204019)
McCombie,W.R.
Direct Submission
Submitted (24-FEB-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-12K1
Center clone name: RP23-12K1
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12288: contig of 12288 bp in length
* 12289: gap of unknown length
* 12370: contig of 12244 bp in length
* 24614: gap of unknown length
* 24694: gap of unknown length
* 24695: contig of 11823 bp in length
* 36518: gap of unknown length
* 36599: contig of 9121 bp in length
* 45720: gap of unknown length
* 45801: contig of 8540 bp in length
* 54341: gap of unknown length
* 54422: contig of 8463 bp in length
* 62885: gap of unknown length
* 62966: contig of 7987 bp in length
* 70953: gap of unknown length
* 71034: contig of 6655 bp in length
* 77689: gap of unknown length
* 77770: contig of 6628 bp in length
* 84398: gap of unknown length
* 84479: contig of 6333 bp in length
* 90812: gap of unknown length
* 90893: contig of 5922 bp in length
* 96815: gap of unknown length
* 96896: contig of 5229 bp in length
* 102125: gap of unknown length
* 102206: contig of 5070 bp in length
* 107276: gap of unknown length
* 107357: contig of 4943 bp in length
* 112300: gap of unknown length
* 112381: contig of 4874 bp in length
* 117255: gap of unknown length
* 117336: contig of 4872 bp in length
* 122208: gap of unknown length
* 122289: contig of 4575 bp in length
* 126864: gap of unknown length
* 131069: contig of 4125 bp in length
```

\* 131070 131150: gap of unknown length  
\* 131151 135196: contig of 4046 bp in length  
\* 135197 135277: gap of unknown length  
\* 135278 139232: contig of 3955 bp in length  
\* 139233 139313: gap of unknown length  
\* 139314 143090: contig of 3777 bp in length  
\* 143091 143171: gap of unknown length  
\* 143172 146888: contig of 3717 bp in length  
\* 146889 146969: gap of unknown length  
\* 146970 150560: contig of 3691 bp in length  
\* 150661 150741: gap of unknown length  
\* 150742 154367: contig of 3626 bp in length  
\* 154368 154448: gap of unknown length  
\* 154449 158070: contig of 3622 bp in length  
\* 158071 158151: gap of unknown length  
\* 158152 161654: contig of 3503 bp in length  
\* 161655 161735: gap of unknown length  
\* 161736 165155: contig of 3420 bp in length  
\* 165156 165236: gap of unknown length  
\* 165237 168655: contig of 3419 bp in length  
\* 168656 168736: gap of unknown length  
\* 168737 172141: contig of 3405 bp in length  
\* 172142 172222: gap of unknown length  
\* 172223 175197: contig of 2975 bp in length  
\* 175198 175278: gap of unknown length  
\* 175279 177884: contig of 2606 bp in length  
\* 177885 177965: gap of unknown length  
\* 177966 180515: contig of 2550 bp in length  
\* 180516 180596: gap of unknown length  
\* 180597 183126: contig of 2530 bp in length  
\* 183127 183207: gap of unknown length  
\* 183208 185685: contig of 2478 bp in length  
\* 185686 185765: gap of unknown length  
\* 185766 188220: contig of 2455 bp in length  
\* 188221 188300: gap of unknown length  
\* 188301 190688: contig of 2388 bp in length  
\* 190689 190768: gap of unknown length  
\* 190769 193133: contig of 2365 bp in length  
\* 193134 193213: gap of unknown length  
\* 193214 195453: contig of 2240 bp in length  
\* 195454 195533: gap of unknown length  
\* 195534 197618: contig of 2085 bp in length  
\* 197619 197698: gap of unknown length  
\* 197699 199717: contig of 2019 bp in length  
\* 199718 199797: gap of unknown length  
\* 199798 201775: contig of 1978 bp in length  
\* 201776 201855: gap of unknown length  
\* 201856 203797: contig of 1942 bp in length  
\* 203798 203877: gap of unknown length  
\* 203878 204019: contig of 142 bp in length.

FEATURES

Source

1. .204019  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-12K1"

BASE COUNT 57429 a 43087 c 42517 g 57444 t 3542 others  
ORIGIN

Query Match 29.3%; Score 747.2; DB 2; Length 204019;  
Best Local Similarity 70.2%; Pred. No. 6.2e-171;  
Matches 1191; Conservative 0; Mismatches 429; Indels 77; Gaps 11;  
QY 175 CCGGTTTCTCGTGGTGAAGGACTCGATAATGTAAGTGGCTCTCCGAAAGACCCG 1628  
Db 1687 CAGGTTTCTCGTGGTGAAGGACTCGATAATGTAAGTGGCTCTCCGAAAGACCCG 1628  
QY 235 CTTGGCCGTAACCTCAGGAAGGTTCTCTTTGACCCCATCTCATTCGAAGCCACCTCTGA 294  
Db 1627 CGTGGCCCTCGCTGAGCAAGATCTCCAAGACCCCTGTCTCGTTTATAGCCACTTGAGA 1568  
QY 295 AGCCACTTGAGAAAAATGATGTGACAGTTCTCTTTGACCCCATCTCATTCGAAGCCACCTCTGA 294  
Db 1627 CGTGGCCCTCGCTGAGCAAGATCTCCAAGACCCCTGTCTCGTTTATAGCCACTTGAGA 1568  
QY 295 AGCCACTTGAGAAAAATGATGTGACAGTTCTCTTTGACCCCATCTCATTCGAAGCCACCTCTGA 294  
Db 1567 AG-----AAATGATGTGACAAATCTTATGAAAAGGACGTTTGTGGAAGTGTAGG 1519

QY 352 CCATCTGTGAAGAAAGTGGCCCTTTCTCCCGCTTGCAAAATAGACATTTCTCAAATTCCAA 411  
Db 1518 ATACCGTGCAGAAGGTCACTCTTCCCTGGTCTGTGAAATAGACATTTCTCAAATTCCAA 1459  
QY 412 AATGCCAGCAAGACCCCAATTTACCTGAAAGCAGCCCAATAACAAGAAAGAAATTT 471  
Db 1458 AATGCTGCCAAGACCCCAATTTACCTGAAAGCGGCTAACACAAGAAAGAAATTT 1399  
QY 472 TAAACTGAGGACATTTCTGCTCCTGATATGATCAGTCCCGCTTGGAGACTTTCCGCCA 531  
Db 1398 TAAACTGAGGACATTTCTGCTCCTGATATGATCAGTCCCGCTTGGAGACTTTCCGCCA 1339  
QY 532 CACCATCCACATTTGGCAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTCTTCTCA 591  
Db 1338 CACTATACACATCGGCAAGAGGGCCAGCATGACGTCTTTGGGATATTTCTCTTCTCA 1279  
QY 592 AGGAACTACGAGCTTTTACCTGGAAACCCAGGAGAGAACACACCTGGGCCAGTCCCTGG 651  
Db 1278 AGGAAATATGAGCTCTTGGCGGGAACCAAGAGAGGACACTCTGGCCAGTTCCTGGG 1219  
QY 652 GCATAATGAGTTCTTCCGGGCCAACAGCACCTCGGACTCAATGTTCACAGAAACACCTC 711  
Db 1218 GCATAACGACTTCTTCCGGGCCAACAGCACCTCGGACTCAATGTTCACAGAAACACCTC 1159  
QY 712 CCGGTGCTCAAAAATGCCATCTCCCTCCCGACCATTTGAGGATCCCAAGCTCTCATGTT 771  
Db 1158 CCGGTGCTCAAAAATGCCATCTCCCTCCCGACCATTTGAGGATCCCAAGCTCTCATGTT 1099  
QY 772 GCGCTTATGTCAACAGTGACATTTAATTTCCAAACAGGAGTCTTTCGGGCCAGCAAGCT 831  
Db 1098 GCGCTTATGTCCCGAGTGACATTTCAATTTCCAAACAGGAGTCTTTCGGGCCCTTAAGTT 1039  
QY 832 GCCAGGCTTAGCTCGAGCCCCGTCTATGGAGGAGAAAGCTCAGGAGAAAGACAGTCTGTT 891  
Db 1038 GCCAAGGCTCAGCTGTGAGCTGTCTATGGAAGAGAAAGTTTCAGGAGCAAGTAGTCTGTT 979  
QY 892 GGAGAAATGGGACAGTCCACAGGGGAGACACCTCGTGGGGCTCCAGCGGTTCTGCATCTCA 951  
Db 978 AGAGAATGGGCGAGTCCACAGGGGAGACACCTCGTGGGGCTCCAGTGGTTCGGGTCTCA 919  
QY 952 GTCCAGCCAAAGGACAGACAGCCACTCTCTCCAGCTGTCCGAACAGTACCCGACTGGCC 1011  
Db 918 GTCCAGCCAGGGCCGACAGCCACTCTCTCCAGCTGTCCGAACAGTACCCGACTGGCC 859  
QY 1012 AGCCGAGGACATGTTTGACCATCCACCCCATGAGGACTCATCAAGGAGAAAGACTAAGTC 1071  
Db 858 CGCGGACGACATGTTTGAGCATCTGCTCCCTCTGAGCTTGTGAAGTCAAGAGCTAAATC 799  
QY 1072 AGAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCTCCCTGCAGCTTGATCTTGGCC 1131  
Db 798 AGAAGAGTCTCTCTGACCTGACCGGCTCCCTGCTCTCCCTGCAGTGGATCTTGGCC 739  
QY 1132 CTCACCTTTGGATGAGTGTCTGAATGTAATGGATAAAATAAGTAACAGATGCCAATTT 1191  
Db 738 TTCACCTTTGGATGAGTGTCTGAATGTCATGGATAAAATAAGTAGCCAGGAGTCAAGCTC 679  
QY 1192 TTTTCTTTGGGTTAAAGGTAC-----AAAAACAACTAACCCACA 1232  
Db 678 TTTTCTTTGGGTTAAGAGGTACCAAAATAAGCAGAAACAAAAATACTAACCCACA 619  
QY 1233 GTTGAAGAGAGGCTTCCGGAGCTG---TATTTGACAGTTTGTGTGGGTTTCTAAA 1288  
Db 618 GCTGATGGAAGAGCTCTGACTGTTTTTATTTGAGCCATGTGATGATGATTTCTAAA 559  
QY 1289 ATAA-----TATTTTACAAAGTATTTTTTACCTGTTATGCCCTGTTTGCAAA----- 1337  
Db 558 ATAACGCTCTCTTCTTACAAAATATTTTAAACCATCATGCCCTGTTTGCAAAATTT 499  
QY 1338 -----AACAAATTTAGAAAAACAAACAAAGCAAAACCTATCTTGGCAAAAAGGAAG 1390  
Db 498 TGAAACAATAATTTTAAATAATTAACAAAAACAAACAAACAAACAAACCTGTTCTGG 439







QY 885 GTCTGTTGAGAAATGGGACACTCCACCAGGAGACACCTCGTGGGGCTCCAGCGGTTCTG 944  
Db 629 GTCTATTAGAAATGGGACACTCCACCAGGAGACACCTCGTGGGGCTCCAGTTCCG 688  
QY 945 CATCTAGTCCAGCCAGGAGAGACAGCCACTCCTCCAGCCTGTCCGAACACAGTACCCCG 1004  
Db 689 GGTCAGTCCAGCCAGGAGGCGGACAGCCACTCCTCCAGCCTGTCCGAACAGTCCCTCCG 748  
QY 1005 ACTGGCCAGCCAGGAGACATGTTGACCATCCACCCCATCCAGCTCATCAAGGGAAGA 1064  
Db 749 ACTGGCCGCGGACGACATGTTGAGCATCCTGCCTCCTGTGAGCTTGTGAAGTCAAGA 808  
QY 1065 CTAAGTCAGAGAGTCCCTCTCTGACCTTACAGGTTCCCTCTCTCCCTGCAGTTGATC 1124  
Db 809 CTAATCAGAAAGAGTCTCTCTGACCTGACCGGCTCCCTGCTCTCCCTGCAGTTGATC 868  
QY 1125 TTGGGCTCCTCACTTTTGGATGAGGTGCTGAATGTAATGGATAAAAAATAAGTAAACAAGATG 1184  
Db 869 TTGGGCTTCACTTTTGGATGAGGTGCTGAATGTAATGGATAAAAAATAAGTAGCAGGAG 928  
QY 1185 CCAACTTTTCTTCTTGGGTAAGAGGTACAAAAACAACTAACACAGTTGAAGA 1240  
Db 929 TCAGCTCTTTTCTTGGGAGTAGAGGTACCAAAATAAGCAGAAAAAATAAAAAA 984

RESULT 13  
G23498/c  
LOCUS human STS WI-13652, sequence tagged site. 432 bp DNA linear STS 31-MAY-1996  
DEFINITION G23498  
ACCESSION G23498  
VERSION G23498.1 GI:1343824  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE Homo sapiens STS derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 432)  
TITLE Hudson, T.  
JOURNAL Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS  
COMMENT Unpublished (1995)

Contact: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu

Primer A: TGCTTGAAGTACCTATGTAATGCA  
Primer B: GATTGTAAAGTGTAAGGCACAGG  
STS size: 131  
PCR Profile:  
Presoak:  
Denaturation:  
Annealing: 56 degrees C  
Polymerization:  
PCR Cycles: 35  
Thermal Cycler:  
Protocol:  
Template: 10 ng  
Primer: each 5 pM  
dNTPs: each 4 nM  
Taq Polymerase: 0.025 units/ul  
Total Vol: 20 ul

Buffer:  
MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 9.3

FEATURES  
source Derived from dbEST (genbank accession R43949).  
Location/Qualifiers  
1..432  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="178.4 CR from top of Chr2 linkage group"  
STS  
33..163  
primer\_bind 33..56  
primer\_bind complement(140..163)  
BASE COUNT 147 a 74 c 64 g 141 t 6 others  
ORIGIN  
Query Match 15.4%; Score 393; DB 11; Length 432;  
Best Local Similarity 96.0%; Pred. No. 5.8e-85;  
Matches 410; Conservative 0; Mismatches 16; Indels 1; Gaps 1;  
QY 1740 CAAAATTTTGGACTGTCTCAG-CTACATGTTGGTAAGACACAGGCAAAAGTACTGTAACTA 1798  
Db 430 CAAAATNTGGACTGTCTCAGCTACATGTTGGTAAGACACAGGCAAAAGTACTGTAACTA 371  
QY 1799 AGTTATTTTAAAGTAAATATATTTTACGTGCCTTTGGCTTTTATTTGAGAGTCTA 1858  
Db 370 AGTTATTTTAAAGTAAATATATTTTAAAGTGCCTTTGGCTTTTATTTGAGAGTCTA 311  
QY 1859 CATTATATAGATTCTACATCAGATGTTGTCACCTATTTCCTATTTGGGATTCCTAAGC 1918  
Db 310 CANTTTATAGATTCTACATCAGATGTTGTCACCTATTTCCTATTTGGGATTCCTAAGC 251  
QY 1919 TGTGTATGTCGCTGTTGGAAAAGTGTATTCATCTAGTTTTCCTTCATCTGTTA 1978  
Db 250 TNGTATGTCGCTGTTGGAAAAGTGTATTCATCTAGTTTTCCTTCATCTGTTA 191  
QY 1979 TCATACTTTTAAACAGCAACCAATAACGGATTGTAAAGTGTAAAGGCACAGGTTACTCATG 2038  
Db 190 TCATACTTTTAAACAGCAACCAATAATGGATTGTAAAGTGTAAAGGCACAGGTTACTCATG 131  
QY 2039 ATGCTTCTGCAGAGACTGTGGGGCTACACCACATATGTTATTTGGAAATATAGGTATTTTA 2098  
Db 130 ATGCTTCTGCAGAGACTGTGGGGCTACACCACATATGTTATTTGGAAATATAGGTATTTTA 71  
QY 2099 GTACAGTACATACTTGCAATTACATAGGTTACTTCAAGCAACACAAATAAAAGTAAATGATA 2158  
Db 70 GTACAGTACATACTTGCAATTACATAGGTTACTTCAAGCAACACAAATAAAAGTAAATGATA 11  
QY 2159 AAGTGAA 2165  
Db 10 AAAAAAA 4

RESULT 14  
AF098290  
LOCUS Homo sapiens Cdc42 effector protein 2 mRNA, complete cds. 741 bp mRNA linear PRI 09-JAN-2001  
DEFINITION AF098290  
ACCESSION AF098290  
VERSION AF098290.1 GI:4323061  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 741)  
TITLE Burchell, P.D., Snow, D.M., Bahou, W. and Spiegel, S.  
JOURNAL MSE55, a Cdc42 effector protein, induces long cellular extensions in fibroblasts  
MEDLINE 99362714  
PUBMED 10430899  
REFERENCE 2 (bases 1 to 741)  
AUTHORS Hirsch, D.S., Pirone, D.M. and Burchell, P.D.  
TITLE A new family of Cdc42 effector proteins, CEPs, function in fibroblast and epithelial cell shape changes  
JOURNAL J. Biol. Chem. 276 (2), 875-883 (2001)



QY 763 TCTCATGTTGCC 775  
      |||||  
Db 441 TCTCACCCCTGCC 453

Search completed: July 25, 2003, 18:27:13  
Job time : 6716.77 secs

